FH Burgenland
UNIVERSITY OF APPLIED SCIENCES
Department Informationstechnologie und Informationsmanagement

Bachelorarbeit 2

Vermeidung menschlichen Kontrollverlustes in Eigendynamik entwickelnden KI-Systemen

zur Erlangung des akademischen Grades
Bachelor of Science in Engineering

Bachelorstudiengang IT Infrastruktur-Management

Eingereicht von: Christian Gossmann
Personenkennzeichen: 1410640014
Datum: 23. Juni 2017

Betreut von: DI Rainer Schmidt

Vorwort

Eine Arbeit zu Künstlicher Intelligenz (KI) ist in heutiger Zeit aktueller denn je. Kaum ein elektronisches Gerät arbeitet noch ohne eigenständige Funktionen. Damit verbunden sind allerdings auch Gefahren des Kontrollverlustes. Spätestens wenn Menschen zu Schaden kommen fühlt man sich an Science- Fiction-Filme aus der Terminator-Reihe erinnert. Mit dieser Erinnerung werden jedoch weder Ursache noch Lösung des Problems autonom agierender Maschinen behandelt. Wie derartige Maschinen das autonome Handeln erlernen, warum man überhaupt Maschinen mit „Intelligenz" ausstattet und welche Möglichkeiten existieren, Herr der Lage zu bleiben, ist Gegenstand dieser Abhandlung.

Die Arbeit soll dabei keinesfalls die Chancen und neuen Möglichkeiten, welche sich durch KIs für die Menschheit bieten, in Abrede stellen. Sie soll allerdings zur Vorsicht mahnen, hier den eigenen Verstand einzusetzen und sich nicht blindlings einer KI anzuvertrauen, besonders dann nicht, wenn das eigene Leben davon abhängig sein könnte.

Bedanken möchte ich mich für die Unterstützung durch die FH-Burgenland, allen voran bei meinem Betreuer DI Rainer Schmidt sowie den vielen Personen, die sich bereits wissenschaftlich mit der Materie auseinandergesetzt haben. Auch meine Familie kam bereits zum zweiten Mal zu kurz, was hiermit wieder gutgemacht werden soll.

Christian Gossmann

Eisenstadt, 13. Juni 2017

Inhaltsverzeichnis

Kurzfassung

In heutiger Zeit werden mehr und mehr elektronische Systeme durch Künstliche Intelligenz gesteuert. Problematisch wird dies dann, wenn infolge derer Fehlfunktionen Menschen zu Schaden kommen. Zu denken ist dabei v.a. an Waffensysteme wie Drohnen und immer weiter entwickelte Roboter, die fast schon Fähigkeiten realer Personen entwickeln. In dieser Abhandlung geht es um das Entstehen der KIs in diesen Systemen und v.a. um die damit verbundenen Gefahren. Hierbei ist an sich entwickelnde Eigendynamik zu denken, die dann nicht mehr kontrollierbar ist. Szenarien wie Unfälle mit autonom fahrenden PKWs oder emotionslosen Drohnen und Robotern, deren Aufgabe der Kampf und damit das Töten von Menschen ist, zeugen davon. Treten v.a. im militärischen Bereich Fehler in den Steuerungssystemen von Waffen auf, können Szenarien eintreten, wie sie in den Terminator-Filmen thematisiert werden, nämlich die Versklavung und Auslöschung der Menschheit.

Anwendung findet KI heute im militärischen Bereich weitaus häufiger als man vielleicht denken mag. Die Tendenz geht hier in Richtung Robotersoldaten und Mikrowaffen, deren Funktion die Ausschaltung gegnerischer Infrastruktur ist. Problematisch dabei ist, wenn kritische Systeme wie eben Waffen außer Kontrolle geraten, auch zwangsläufig Menschenleben gefährdet sind. Selbiges gilt für den Zivilbereich, wenn autonome Fahrzeuge Unfälle verursachen. In vorliegender Abhandlung soll den Ursachen für derartige Fehlfunktionen nachgegangen werden. Von besonderem Interesse ist dabei, welche Möglichkeiten existieren, diesen Kontrollverlust über Maschinen zu vermeiden. Es werden dabei die Grundlagen von KIs behandelt, was dabei eine KI von klassischer Software unterscheidet und wo Science-Fiction bereits Realität geworden ist. Aufgezeigt werden entsprechende Strategien, wie Maschinen überhaupt lernen, was sie konkret tun sollen. Hinsichtlich der Ursachen für Fehlverhalten werden dabei die immer wieder gemachten Fehler im Entwicklungsprozess von Software beleuchtet.

KIs sind nämlich nichts anders als von Menschen geschaffene Programme. Wer bereits mit Betriebssystemen oder sonstiger Anwendungssoftware gearbeitet hat, wird möglicherweise auch wissen, wie zuverlässig diese Programme tatsächlich sind – nämlich gar nicht. Wer hier noch nie erlebt hat, wie eine Anwendung sich unvorhergesehen verhält, anstatt wie man es eigentlich von ihr erwartet, lügt entweder oder hat die letzten Jahrzehnte der digitalen Revolution verschlafen. Warum also sollten sich KIs hier anders verhalten als Standardsoftware?

Abstract

This study examines how artificial intelligence in our days controls more and more electronic systems. This becomes problematic when human beings are harmed as a result of malfunctions. Development of weapon systems such as drones and robotics is advancing faster and faster. This leads to abilities that even real people have. With the emergence of the AIs in these systems and the dangers arising from them, scenarios might become true, like the Terminator movies are showing to us. In this case it is necessary to think about strategies in preventing such an evolving AI, which can no longer be controlled. Scenarios such as accidents with autonomous cars or emotion-free drones and robots, whose task is the fight and killing of humans is evidence of this.

Today, AI already is much more used by military forces than being thought about. State of the art are robot soldiers and microweapons, whose function is the annihilation of opposing infrastructure. However, the problem is, when critical systems such as weapons are out of control, innocent human lives are endangered. The same applies to the civil sector, especially, when autonomous vehicles cause accidents. In this paper the causes of such malfunctions will be investigated. This will be shown in detail in the chapter about current AI-Systems, while the following chapters will focus on how to avoid loss of control over AI-driven machines. What an AI distinguishes from classical software will be part of the study. Where science fiction already has become reality will be shown too, as well as Strategies, how machines will learn what they are supposed to do. The goal is, to show how errors and their causes in the development process of control software for AI can be avoided.

AI is nothing other than a human-created program. Anyone who already has worked with operating systems or other application software might also know how reliable these programs really are – not at all. Who has never experienced unpredictable behavior in applications, either is a liar or has been living behind the moon during, while the digital revolution took place. So why should KIs behave differently from other software? Finally the conclusion gives specific advice in how to prevent loss of control over autonomous machines.

1 Einleitung

1.1 Problemstellung

"Proponents of fully autonomous weapons suggest that the absence of human emotions is a key advantage, yet they fail adequately to consider the downsides. Proponents emphasize, for example, that robots are immune from emotional factors, such as fear and rage, that can cloud judgment, distract humans from their military missions, or lead to attacks on civilians. They also note that robots can be programmed to act without concern for their own survival and thus can sacrifice themselves for a mission without reservations." (Docherty, 2012, S.37)

Dieses Zitat des Human Rights Watch-Berichtes zu autonomen Waffen ist rund 30 Jahre nach Beginn der Terminator-Filmreihe im Jahr 1984 Realität geworden. Immer häufiger werden Bereiche menschlichen Lebens autonom handelnden Maschinen anvertraut. Omnipräsent sind interaktionsfähige Roboter v.a. in Japan, die kaum von Menschen zu unterscheiden sind und bis 2020 mit Menschen zusammen leben sollen. (Mainzer, 2016, S.142). Dazu kommen selbstständig fahrende Autos und fliegende Drohen für militärische oder Überwachungszwecke (List, 2017, S.30ff.). Gesteuert werden alle diese Systeme durch entsprechende Algorithmen, bzw. Künstliche Intelligenzen. Der weltweit renommierte Physiker Stephen Hawking erblickt in dieser Entwicklung, die den Menschen übertrifft, die Gefahr des Endes der Menschheit. Dabei geht er von zwei verschiedenen KIs aus. Erstens wären primitive zu nennen, wie es sie heute z.B. in verschiedensten elektronischen Geräten des Internet of Things gibt, wobei diese durchaus für den Menschen nützlich sein mögen. Dies gilt allerdings nur solange, solange er auch die Kontrolle darüber behält. Zweitens gibt es jene KIs, die infolge ihrer Geschwindigkeit die biologische Evolution übertreffen. Ein Mensch kann eben nicht in Teraflops rechnen und denken.

Eine Demonstration derartiger KIs, die besser als Menschen sind, wurde der breiten Masse 1997 vor Augen geführt, als der Schachcomputer „Deep Blue" den Weltmeister Garry Kasparov besiegte. Einwenden könnte man an dieser Stelle natürlich, dass KIs nur Algorithmen folgen und gesetzt den Fall keiner vorliegender physischen Defekte an der Hardware auch keine Fehler machen. Menschen wie Kasparov unterliegen dagegen tagesverfassungsabhängigen Schwankungen in der Konzentration und einem Stressniveau. Computer jedoch empfinden weder Stress noch Konzentrationsschwächen und haben keinerlei Emotionen, wie im einleitenden Zitat angeführt. Weniger bekannt ist allerdings der tatsächliche Umstand, der zum Sieg über Kasparov führte. Es lag ein Steuerungsfehler in der KI von Deep Blue vor. Wie einer der Programmierer 2012 zugab, war Deep Blue beim 44. Zug der entscheidenden Partie nicht in der Lage, einen Gegenzug zu errechnen. Es wurde daher ein Notfall-Fallback auf einen zwar gültigen, aber per Zufall gewählten Zug vollzogen. Dieser wirkte amateurhaft, verwirrte Kasparov letztlich allerdings dermaßen, dass er selbst

einen Fehler beging und damit Deep Blue den Sieg ermöglichte (Silver, 2012, S.233ff.). Es war 1997 also noch pures Glück und keine taktische Künstliche Intelligenz. Die Gefahr dieser Entwicklung liegt allerdings auf der Hand und zeigt, wie die Entwicklung sehr wohl in Richtung der Handlung der Terminator-Filme geht. In der Serie Terminator SCC wurde auch nicht zum Spaß eine Schach-KI namens „der Türke" als Ursprung der die Menschen auszurotten versuchenden Superintelligenz Skynet gewählt.

Mit dem Einzug elektronischer Steuersysteme, bzw. KIs in quasi alle Lebensbereiche stellt das eigentliche Problem dabei entweder deren Komplettversagen oder auch nur Teilversagen durch fehlerhaftes Implementieren von Verhaltensweisen dar. Es drohen dabei finanzielle Schäden oder Gefahr für Leib und Leben. Konkrete Beispiele für finanzielle Einbußen durch die bei Hawking genannten primitiven KIs infolge von Fehlern wären z.B. unkontrollierbare Datenübertragungen „intelligenter" Kühlschränke (Putnik, 2017, S.62), die selbstständig Lebensmitteln bestellen, welche natürlich auch bezahlt werden müssen. Der schon als Klassiker einzustufende Fehler schlechthin ist allerdings der Unterlauf eines fälschlicherweise vorzeichenlos gewählten Integers. Hier hatte ein vermeintlicher Glückspilz in einem Spielcasino eigentlich sein Geld verloren, allerdings trotzdem gewonnen. Wie kommt dieser Widerspruch in sich zustande? Die Lösung liegt in der binären Wertinterpretation begründet. Negative Zahlen konnte das System beim vorzeichenlosen Integer nicht darstellen. Stattdessen erzeugte der binäre Wert Null vermindert um eins bei zwei hoch 32 gesetzten Bits einen positiven Wert und einen „Gewinn" von rund 43 Millionen €. Es folgten zwei Jahre Prozess und letztlich eine gütliche Einigung auf eine Million € als Gewinnsumme (Jelenko, 2012) – ein teurer Fehler also für das Casino. Auch Bestellungen durch vermeintlichen Zuruf und Datenübertragungen aus dem privaten Lebensbereich zu Serven im Internet wie z.B. aktuell bei Amazon Echo, fallen in diese Kategorie. (Hosbach, 2017, S.42).

Während sich diese Beispiele harmlos anhören, sind die dabei gemachten Fehler letztlich auch die Quelle der Fehler in autonomen Waffensystemen und autonomen Maschinen. Diese Maschinen müssen dabei nicht einmal Waffen sein, um zur Waffe zu werden. Im Straßenverkehr sorgte erst kürzlich ein Unfall für Schlagzeilen, wo ein per Fahrassistenzsystem gesteuertes Fahrzeug kein Bremsmanöver durchführte als ein Hindernis in Form eines LKWs mit weißer Bordwand auftauchte. Das Ergebnis war, dass der Lenker infolge des ungebremsten Aufpralls verstarb. Untersuchungen ergaben, dass das Steuerungssystem, also die KI, fälschlicherweise davon ausging, bei der weißen LKW-Bordwand hätte es sich um einen über der Fahrbahn hängenden Wegweiser gehandelt. Aus Sicht der KI lag somit keinerlei Fehler vor – es bestand ja keine Gefahr. Allerdings hat es sich bei der KI um ein reines Assistenzsystem gehandelt und nicht um einen echten Autopiloten. Insofern war der Hersteller frei von Schuld, da entgegen der Betriebsanleitung der

Fahrer die ausschließliche Steuerung dem Assistenzsystem überließ, wie erst kürzlich gerichtlich festgestellt wurde (Kotrba, 2017).

Leider ist dieses Beispiel mit autonom fahrenden Fahrzeugen kein Einzelfall. Todesfälle treten heute potentiell immer dort auf, wo KIs am Werke sind und deren Programmierung falsche Reaktionen auslösen. So tötete unlängst ein Industrieroboter eine Frau, die mit Wartungsarbeiten beauftragt war (Kronen Zeitung, 2017, S. 29). Wenngleich hierüber die Ursache noch Spekulation ist, sicher ist auch hier, dass man KI-Systemen nicht blindlings vertrauen darf. Hawking unterstellt superintelligenten KIs dabei nicht einmal Bosheit, sondern im Gegenteil, die Fähigkeit der KI selbst. Diese vergleicht er mit einem Betreiber eines hydroelektrischen Projektes. Auch wenn keine mutwillige Zerstörung eines Ameisenhaufens in einem zu flutenden Gebiet erfolgt, geschieht dies trotzdem in gutgläubiger Ausführung (Költzsch, 2015). Umgemünzt auf die oben genannten autonomen Waffensysteme kann man Hawking nur zustimmen, denn diese töten ebenfalls in gutem Glauben auf Anweisung der jeweiligen Regierungen Terroristen, bzw. zerstören völkerrechtlich legitime, militärische Ziele in Konfliktfällen - die Realität sieht jedoch anders aus, da in den Medien zu oft von Kollateralschäden in Form getöteter Zivilisten und Kinder, bzw. fehlgeleiteten Treffern auf Krankenhäuser oder sonstige Infrastruktur berichtet wird.

Hinter jeder dieser KI steckt letztlich Programmierung durch Algorithmen, die wiederum - derzeit noch - von Menschen entwickelt werden. Hier müsste für alle auftretenden Eventualitäten Vorsorge getroffen werden, um Fehler auszuschließen, was faktisch ein Ding der Unmöglichkeit ist. Wenn man sich den Begriff ansieht, „fehlt“ also etwas. Daher tritt maschinelles Lernen in Kombination mit Sensoren und Speicherbänken in Aktion. Schon 1986 finden sich bei Heinemann entsprechende Erläuterungen, wie Roboter diesbezüglich funktionieren. Primitive Roboter sind hardwaremäßig kodiert. D.h. konkret, dass sie ihr verlötetes „Programm“ abarbeiten und z.B. Werkstücke für eine Schweißmontage an exakter Koordinatenposition erwarten. Abweichungen führen hierbei zu Fehlfunktionen. Programmierbare Roboter dagegen „lernen“ aufgrund ihrer Sensoren, Werkstücke, die ein wenig verschoben sind, zu erkennen und passen ihr Schweißprogramm entsprechend den geänderten Koordinaten an (Heinemann, 1986, S.13). Es kann für obigen Fall der getöteten Frau nur vermutet werden, dass auch hier eine Fehlinterpretation der Sensoren vorlag oder es dem Roboter generell an einer KI mangelte, indem er hart kodiert war und irrtümlich aktiviert wurde. Denkbar wäre auch, wie vorhin angeführt, ein Kollateralschaden, falls der Roboter in gutem Glauben geschweißt hat. Der Kontrollverlust bezieht sich dabei v.a. auf das Fehlen von Sicherheitseinstellungen und Abbruchbedingungen sowie auf die entwickelte Eigendynamik der Maschinen, die erst durch Programmfehler entsteht. Prof. Taschner von der TU Wien sähe es z.B. zwar gern, wenn Maschinen bei Aufgaben, die sich ausschließlich durch Algorithmen lösen lassen, auch die

entsprechende Entscheidungen selbstständig treffen würden. Wo dagegen Entscheidungen anstehen, die über rein logisches Denken hinausgehen, sollten KIs diese nicht treffen (nach Strackel, 2017, S.28).

1.2 Zielsetzung

Zu obigen Problemen stellt sich daher folgende Forschungsfrage: Wie wird Kontrollverlust menschlichen Handelns in KI-Systemen vermieden und sichergestellt, dass Fehler so gut wie ausgeschlossen werden? Hundertprozentige Sicherheit wird es zwar nie geben können, allerdings geht es darum, zumindest diejenigen Fehler auszuschließen, die eines Tages dazu führen können, die Maschinen tatsächlich die vollständige Kontrolle übernehmen zu lassen, wie Mainzer diese Gedanken aufwirft (Mainzer, 2016, S.1ff.) und die Terminator-Filme dies treffend vor Augen führen.

Um die Forschungsfrage zu beantworten ist es Ziel, Strategien aufzuzeigen, die letzte Entscheidungsgewalt über immer intelligenter werdende Geräte zu behalten. Ferner soll der Begriff des KI-Systems behandelt werden, wobei hier im Speziellen von Interesse ist, wie der eigentliche Lernprozess in Maschinen stattfindet. So kann man nämlich bei einem klassischen Kaffeeautomaten nicht davon ausgehen, dass dessen Steuerlogik ein KI System darstellt, folgt er doch hart kodierten elementaren Gesetzen der Schaltalgebra (Rechenberg, 1997, S.82ff.). Dort wo jedoch KIs zum Einsatz kommen, also komplexere Datenstrukturen und Programme vorliegen, liegt das Hauptaugenmerk auf den Möglichkeiten diese fehlerfrei zu gestalten und gegen unvorhergesehenes Verhalten zu sichern. Alan Turing zufolge ist eine Maschine nämlich dann als intelligent anzusehen, wenn eine Versuchsperson jeweils einem menschlichen und maschinellen Partner versteckt gegenübersitzt und beide anhand ihres Verhaltens nicht voneinander unterscheiden kann (Dorn, Gottlob, 1997, S.820). Einwenden könnte man natürlich, dass auch der Kaffeeautomat intelligent ist, wenn letztlich sowohl ein Mensch als auch der Automat gleichwertigen Kaffee herstellen würden. Obiger Automat ist allerdings nicht interaktionsfähig, um konkret einen bestimmten Menschen zu identifizieren und danach die gewünschte, da z.B. erlernte Menge an Zucker oder Milch in den Kaffee zu geben, damit man ihn zumindest potentiell für einen Menschen halten kann. Dies könnte sich aber durch entsprechende Sensoren, Spracherkennung und Variationen in der Getränkezubereitung ändern. Dann allerdings müsste der Automat auch komplexere Datenstrukturen und damit erheblich mehr Schaltungen analog den Neuronen in einem menschlichen Gehirn aufweisen.

Sicherung elektronischer Systeme mit klassischen Strategien wie Antiviren-Schutz, Einsatz von Firewalls oder Backups als Teilbereich des Sicherheitsmanagements, eine KIs selbst vor Hacking-Angriffen zu schützen, ist dagegen nicht Gegenstand der Betrachtung. Ferner ist es nicht Ziel Datenschutz-, bzw. rechtliche Aspekte zu behandeln, ebenso wenig wie gezieltes Data-Mining, da dies mehr in den Bereich PR und Marketing fällt.

2 Grundlagen

2.1 Definition KI

Für den Untersuchungsgegenstand KI ist eine eindeutige Definition unerlässlich. In der Literatur finden sich diesbezüglich unterschiedliche Definitionen. Dies zeugt daher von einem Fachgebiet, in das unterschiedliche wissenschaftliche Disziplinen einfließen. So haben alleine Russel und Norvig acht Definitionen von KI zusammengetragen, nämlich:

bezogen auf menschliches Denken einerseits

- Computern das Denken beizubringen und andererseits
- Automatisierung von Aktivitäten, die wir dem menschlichen Denken zuordnen

bezogen auf rationales Denken die

- Studie mentaler Fähigkeiten durch die Nutzung programmiertechnischer Modelle sowie das
- Studium derjenigen mathematischen Formalismen, die es ermöglichen, wahrzunehmen, logisch zu schließen und agieren

bezogen auf menschliches Handeln

- die Kunst, Maschinen zu schaffen, die Funktionen erfüllen, die, werden sie von Menschen ausgeführt, der Intelligenz bedürfen und
- das Studium des Problems, Computer dazu zu bringen, Dinge zu tun, bei denen ihnen momentan der Mensch noch überlegen ist

bezogen auf rationales Handeln, dass

- Computerintelligenz hierbei die Studie des Entwurfs intelligenter Agenten darstellt und sich
- KI mit intelligentem Verhalten in künstlichen Maschinen beschäftigt (Russel, Norvig, 2012, S.23)

Anhand obiger Definitionen von KIs lässt sich die Erkenntnis gewinnen, wonach es sich dabei um Nachbildung menschlichen Denkens in rationaler Form, das Maschinen beigebracht wird, handelt. Ziel ist es somit, Probleme zu bewältigen, die momentan nur der Mensch erfüllen kann. Dazu bedarf es externer Agenten im Sinne von Sensoren und der Kenntnis über das Verhalten von Maschinen, konkret also deren Hard- und softwaretechnischen Aufbau. Für KI selbst bringt Mainzer dagegen folgende Arbeitsdefinition: „Ein System heißt intelligent, wenn es selbstständig und effizient Probleme lösen kann. Der Grad der Intelligenz hängt von der Selbstständigkeit, dem Grad der Komplexität des Problems und dem Grad der Effizienz des Problemlösungsverfahrens ab.“ (Mainzer, 2016, S.3)

Kennzeichnend für die Selbstständigkeit von KIs sind dabei sog. Agenten. Agenten werden z.B. in Software dazu verwendet, Aktionen zu setzen. So gibt es z.B. Agenten, die auf Servern installiert werden und dafür zuständig sind mit einer zentralen Steuerung zu kommunizieren. Ziel hierbei kann es z.B. sein, zu sichernde Dateien zur Verfügung zu stellen, die auf zentralen Server gesichert werden (z.B Backup Exec). Auch System-Inventarisierung wäre denkbar (z.B. Deskcenter) oder Fernsteuerung des Systems schlechthin (z.B. UVNC). Diese klassischen Agenten sind vereinfacht ausgedrückt ebenfalls nur Software. KI-Agenten dagegen sind auf rationales Handeln ausgelegt, d.h. das beste Ergebnis zu erzielen oder im Falle von Unsicherheiten zumindest das beste erwartete Ergebnis. Einem klassischen Backup Agenten ist es dagegen egal, momentan offene Dateien nicht sichern zu können, ebenso wie bei Inventarisierungen ein infolge von Netzwerkproblemen nicht verfügbares System nicht erfasst werden kann. Selbiges gilt für die Fernwartung. KIs dagegen könnten sich „merken“, dort und da auf Schwierigkeiten gestoßen zu sein und nach definierten Zeitspannen erneut versuchen, die fehlerhafte Handlung nachzuholen. Diese Aktion müsste allerdings programmtechnisch festgelegt werden, d.h. konkret, auch rationales Handeln erst einmal nach entsprechenden Denkregeln abzubilden. Man merkt, je komplexer die Entscheidungsfindungen von sog. Wenn/Dann-Strukturen wird, hier alle möglichen Szenarien abzudecken, sich auch die Komplexität einer KI erhöht. Konsequenterweise gehen Russel und Norvig davon aus, dass die perfekte Rationalität - immer das Richtige zu tun - in komplexen Umgebungen mit heutigen Computerleistungen noch nicht erreichbar ist. Agenten selbst erhalten ihren Input durch Sensoren und setzen mittels Aktuatoren Handlungen (Russel, Norvig, 2012, S.25ff). Bezogen auf den einleitenden Schweißroboter wäre dabei an optische Sensoren zu denken, die ein Werkstück erfassen und die Schweißarme durch Aktuatoren entsprechend steuern. Die Gefahr besteht nun aber nicht im Agenten, sondern in der Kombination aus den Eingangsdaten, bzw. in resultierenden Handlungen. Nicht der Agent kann töten, sondern die Kombination aus nicht erreichbarer perfekten Rationalität in Verbindung mit fehlerhaften Eingangsdaten von Sensoren und entsprechend gesteuerten Aktuatoren. Damit ist auch der Weg vorgezeichnet, wonach heutige KIs immer potentielle Fehlerquellen beherbergen. Genau damit sind die Fehler gemeint, die Menschenleben kosten können, wenn die Systeme in kritischen Umgebungen Anwendung finden. Kritische Umgebungen trifft man allerdings mit der fortschreitenden Digitalisierung und im Zeitalter des Internet of Things immer häufiger an, sowohl im zivilen wie im militärischen Bereich.

2.2 Die Geburtsstunde der KI und aktueller Trend

Gemeinhin wird als Geburtsstunde der KI Alan Turings Abhandlung über „Computing Machinery and Intelligence“ betrachtet. Darin erläutert Turing das Entstehen von Intelligenzen und begründet dies mit digitalen Computern im Sinne frei programmierbarer Systeme. Als Hauptproblem wird eben jene

Programmierung betrachtet. Damals schon dachte Turing an die Nachbildung menschlicher Denkprozesse und wusste um die Beschränkung damaliger Hardware Bescheid. 1950 war Hardware nämlich nicht im Entferntesten so leistungsfähig wie heute. Konsequenterweise räumte Turing damals ein:

„As I have explained, the problem is mainly one of programming. Advances in engineering will have to be made too, but it seems unlikely that these will not be adequate for the requirements. Estimates of the storage capacity of the brain vary from 10^{10} to 10^{15} binary digits. I incline to the lower values and believe that only a very small fraction is used for the higher types of thinking" (Turing, o.D., S. 451).

KIs sind heute nichts anderes als genau diejenigen Softwareprodukte, von denen Turing ausging. Damit sind KIs genauso fehleranfällig wie Standardsoftware. KI-Software ist allerdings vielfach komplexer und auftretende Fehler darin folgenreicher. Wie Bostrom einräumt, wird das Ziel, wonach sich Software so verhält, wie sie sich eigentlich verhalten soll, schon bei Standardsoftware sehr oft verfehlt. Service Packs, Updates und Patches zeugen davon. Tritt in heimischen Computern ein Fehler auf mag dies zwar ärgerlich sein, die Welt geht davon jedoch in der Regel nicht unter. Tritt dagegen ein Fehler im Steuerungssystem für Kernkraftwerke oder einem Abschusssystem für Nuklearwaffen auf, sieht die Lage schon gänzlich anders aus. Bostrom denkt hier an Möglichkeiten, die Selbstverbesserung einer sog. Saat-KI bietet:

„Excel hat keine Subroutine, die insgeheim nach der Weltherrschaft strebt und bloß zu dumm dazu ist. Eine Tabellenkalkulation will überhaupt nichts, sie führt einfach blind Instruktionen aus. Was (so könnte man sich fragen) hindert uns daran, eine Anwendung der gleichen Form zu entwickeln, die etwas mehr allgemeine Intelligenz besitzt? Das könnte zum Beispiel ein Orakel sein, das auf eine Beschreibung eines Ziels mit einem Plan für dessen Realisierung antwortet - genau so, wie Excel auf eine Spalte voller Zahlen mit der Berechnung der Summe dieser Zahlen reagiert. Beide drücken damit weder irgendwelche „Präferenzen" in Bezug auf den Output noch in Bezug auf das aus, was Menschen damit tun werden." (Bostrom, 2014, S.215f.)

In den 80er-Jahren des vergangenen Jahrhunderts geisterte diesbezüglich der Begriff der Expertensysteme durch die EDV-Welt. „Expertensysteme sollen die Tätigkeit von (menschlichen) Experten unterstützen bzw. teilweise mechanisieren. Expertensysteme werden also durch das bestimmt, was Experten machen." (Raulefs, 1982, S.62). Dabei gehört es zu deren Aufgabe zu interpretieren, diagnostizieren, planen, konstruieren, beweisen und im Sinne von Tutoring Wissen zu vermitteln (Ebd., S. 62).

Heute ist die Richtung dagegen klar vorgegeben hin zu neuronalen Netzen, die menschliche Gehirne nachbilden und Verknüpfungen herstellen können, ganz wie es ein Mensch zwischen einzelnen Fachgebieten tut. Zwar hatten neuronale Netze schon wesentlich früher ihre Anfänge, als 1951 Marvin Minsky den Neurorechner Snarc entwickelte, sich durchzusetzen beginnen derartige

Systeme aber erst mit entsprechend leistungsfähiger Hardware von heute. Heutige Hardware ist nämlich in der Lage, menschliche Gehirne zu emulieren. Gehirn-Computer-Schnittstellen sollen hier primär eigentlich der medizinischen Anwendung in der Behandlung von Depressionen und Parkinson dienen (Clausen, 2015, S.74ff.), der Weg ist allerdings schon vorgezeichnet zu Kopiermöglichkeiten von menschlichen Gehirnen. Bisher dienen die Schnittstellen noch dazu Steuerungen und Kommunikation mit Versehrten, die auf den Rollstuhl angewiesen sind zu ermöglichen. In Experimenten mit Rhesusaffen ist es allerdings bereits gelungen, diese einen Rollstuhl durch pure Gedankenkraft zielgerichtet steuern zu lassen. (Eberl, 2016, S.348).

Einblicke in sich abzeichnende Eigendynamiken erhält man aktuell z.B. mit dem Betriebssystem Windows 10. Seit dieser Version ist es nicht mehr in vollem Umfang möglich, Updates, die Hersteller Microsoft vorsieht, abzulehnen. Zwar ist es noch möglich, den entsprechenden Update-Dienst manuell zu deaktivieren, per Default-Einstellung werden jedoch in den Home-Versionen sämtliche Updates durchgeführt und zwar im Hintergrund ohne Wissen der Anwender. In den Professional-Versionen kann man diese lediglich hinauszögern. Problematisch in allen Fällen erscheint, es hier mit einem System zu tun zu haben, welches ungewollt neu booten kann. Ist man gerade mit wichtiger Arbeit beschäftigt und hat womöglich nicht rechtzeitig seine Daten abgespeichert, kann dieses Verhalten zum Verlust selbiger führen.

2.3 benachbarte Disziplinen

KIs selbst sind nicht ausschließlich das Produkt reiner Computerwissenschaft und Ingenieurskunst, sondern es spielt auch die Philosophie eine zentrale Rolle, v.a. hinsichtlich der Fragen, wie formale Regeln verwendet werden können, um gültige Schlüsse zu ziehen. Ferner gilt es zu klären, wie aus einem physischen Gehirn mentaler Verstand entsteht, woher Wissen stammt und wie Wissen zu einer Aktion führt. Aus philosophischer Sicht mag wohl die Frage des Verstandes von zentraler Bedeutung sein. Blaise Pascals „Pascaline“ - Rechenmaschine von 1642 schien hier dem menschlichen Verstand näher zu kommen, als alles, was Tiere zu leisten im Stande sind. Thomas Hobbes beschrieb ferner in seinem Leviathan 1651 die Idee eines künstlichen Tieres und verglich dessen Herz als eine aufziehbare Feder, Nerven als Stränge und Räder als Gelenke. René Descartes brachte das Problem ebenfalls auf den Punkt und sprach künstlichem Verstand den freien Willen zur Entscheidung ab: „Wenn der Verstand völlig von logischen Regeln gesteuert wird, hat er auch nicht mehr freien Willen als ein Stein, der 'entscheidet', zum Mittelpunkt der Erde zu fallen.“ (nach Russel, Norvig, 2012, S.27)

Diese Ansätze spiegeln die gesamte Problematik einer KI wider. KIs sind von Menschenhand geschaffen, genau wie Rechenmaschinen und Federn, Stränge, bzw. Gelenken damals. Heute sind dies eben nur Silizium-Chips, Sensoren und Aktuatoren.

Auch Mathematik hinsichtlich Logik, Berechenbarkeit, Algorithmen und Wahrscheinlichkeiten - alles in allem also elementare Dinge, die der Philosophie am Weg zu einer formalen Wissenschaft fehlten, fließen hier mit ein. Zudem gehört Wirtschaftswissenschaft hinsichtlich des Nutzwertes, die Neurowissenschaft aus der medizinischen Forschung, die Psychologie hinsichtlich kognitiver Wahrnehmungen als Informationsinput und letztlich die technische Informatik als entscheidende Disziplin zur eigentlichen Herstellung entsprechender Computer, die KIs beherbergen, dazu. Russel und Norvig erwähnen am Rande auch die Kybernetik und Linguistik. Dies muss vor dem Hintergrund gesehen werden, hier letztlich auch Robotern nicht anzumerken, dass sie KIs darstellen und daher nun mal über linguistische Fähigkeiten verfügen müssen (Russel, Norvig, 2012, S.27ff.)

2.4 aktueller Forschungsstand

Momentaner Forschungsstand ist unter Einbeziehung vorhin genannter Disziplinen das Scheitern sämtlicher Definitionen an der Begrifflichkeit der „Intelligenz“. Was „künstlich“ bedeutet, ist dagegen eindeutig - nämlich nicht von der Natur geschaffen. Was dagegen Intelligenz darstellen soll, ist in der Literatur bis dato strittig und nicht eindeutig geklärt. Warum sonst sollten wie o.a. Russel und Norvig derartig viele Definitionsversuche wagen? In dieser Arbeit wird allerdings der Begriff der KI rein im technischen Sinne verwendet, konkret also Nachbildung menschlicher Denkmuster im Sinne von Verknüpfungen von einzelnen Wissensgebieten. Wie so viele Dinge im Leben dienen diese neben zivilen Nutzungsmöglichkeiten auch militärischen Zwecken. Konkret interessieren sich daher staatliche Behörden – allen voran die Atommächte im UNO-Sicherheitsrat - wie sich KIs zu genau diesem Zweck nutzen lassen. Was dabei technisch machbar ist, bringt Tuck treffend auf den Punkt:

„Wie bei den fliegenden Robotern der Luftwaffe wären auch Roboter zu Wasser technisch in der Lage, den Schießbefehl eigenständig zu geben. Laut Hersteller sind sie heute schon in der Lage, Freund von Feind zu unterscheiden, womöglich präziser als der menschliche Schütze. Zurzeit ist diese Entscheidung per Gesetz den Menschen vorbehalten. Killer-Roboter gibt es zu Land, zu Wasser, in der Luft. Sie verfügen über Künstliche Intelligenz und handeln weitgehend autark. Künstliche Intelligenz wäre jederzeit in der Lage, koordiniert Attacken völlig autark durchzuführen. Das ist Stand der Technik.“ (2016, S.89)

Anzumerken wäre hierzu, es mit einem Gleichgewicht der Kräfte zu tun zu haben. Staatliche Führer der sog. Atommächte wissen hier sehr wohl um den Umstand Bescheid, als erster zu schießen, dann allerdings als zweiter zu sterben. Dies hält somit seit den Zeiten des Kalten Krieges die betreffenden Personen davon ab, aufeinander mit Nuklearwaffen loszugehen. Wer jedoch eine KI davon abhalten soll, wenn diese die Vernichtung der Menschen „beschlossen hat“, diese Antwort kennt derzeit noch niemand.

3 Methodik und persönliches Vorgehen

Zur Beantwortung der Forschungsfrage wird eine systematische Literaturrecherche durchgeführt, wobei der Schwerpunkt auf der Verknüpfung zentraler Begriffe wie „Künstliche Intelligenz“, „Steuerung“, „Fehlerbehebung“, und „Maschinen“ liegt, die auch für die Recherche in Bibliotheken und Online-Katalogen verwendet werden.

Im Kapitel zur Entstehung von KIs werden historische Entwicklungen aufgezeigt, die zur Entstehung von KIs führen. Dabei wird auf die Schwierigkeiten eingegangen, es hier mit einem interdisziplinären Forschungsfeld zu tun zu haben. Neben der eigentlichen Informationstechnik fließen dabei u.a. auch o.a. philosophische, religiöse, psychologische, wirtschaftswissenschaftliche, etc… Aspekte ein.

Im Anschluss daran erfolgt im Kapitel zur Anwendung von KIs im Alltag die Konfrontation des Lesers mit aktuellen Möglichkeiten und Gefahren heute existierender Systeme. Der Schwerpunkt liegt dabei im militärischen Bereich, aber auch in einigen zivilen Anwendungsmöglichkeiten.

Das Kapitel bezüglich Lernens von Maschinen gibt einen Überblick über die Schwierigkeiten, einer Maschine, bzw. KI den menschlichen Verstand und so einfache Dinge wie den Duft einer Rose zu veranschaulichen. Exemplarisch wird auf Schach, Tic Tac Toe und Go eingegangen.

Das entscheidende Kapitel zur Fehlervermeidung in KIs erläutert schließlich die Ursachen, warum KIs außer Kontrolle geraten können und welche Gegenmaßnahmen zu treffen wären. Dabei wird auf immer wieder gemachte Fehler in der Entwicklung von Software eingegangen, denn KIs basieren ja, wie bereits erläutert wurde, darauf.

Zusammenfassend werden dann sämtliche Aspekte bewertet und die Forschungsfrage beantwortet, wie denn der Kontrollverlust des Menschen über Maschinen verhindert werden kann.

4 Entstehung einer KI

Beim Entstehen einer KI gilt es, deren Einsatzzweck im Auge zu behalten. Kennzeichnend ist entweder eines oder mehrere Endziele, auf die hingearbeitet wird. Bostrom spricht in diesem Zusammenhang von einer Saat-KI. Was diese anbelangt, ist allerdings zweifelhaft, welche Endziele man dieser mitgibt. Soll diese Ratschläge der Programmierer annehmen, soll ihr nur ein Endziel eingegeben werden, soll ihr als Endziel die Akzeptanz eines anderen Endzieles ermöglicht werden? Fest steht, in allen Fällen eine Fähigkeitenkontrolle implementieren zu müssen, um der Saat-KI nicht die Kontrollübernahme zu ermöglichen (Bostrom, 2014, S.269). Gemeint ist hier selbstverständlich die Kontrolle über die Menschheit.

Eine allumfassende, sämtliches Wissen beinhaltende KI zu erzeugen schwebte auch den Google-Gründen Larry Page und Sergey Brian vor. Ihnen ging es hier eigentlich gar nicht um eine Suchmaschine, sondern darum, die Welt zu verbessern. Freilich handelt es sich dabei um ein Standardargument, denn natürlich müssen auch Page und Brian von etwas leben. Somit standen sehr wohl kommerzielle Interessen im Vordergrund. Auch US-Geheimdienste waren maßgeblich daran interessiert und betrieben Sponsoring dafür (Schlieter, 2015, S.19). Immerhin lassen sich schon heute mit dem gespeicherten Wissen von Google und der Kombination aus diesen Daten hervorragend Persönlichkeitsprofile von einzelnen Individuen bilden, die sich dann optimal zur Terrorabwehr und zur Planung von Kampfeinsätzen nutzen ließen.

4.1 Der Mensch als genetisches Vorbild

V.a. das Entschlüsseln der Gene und das Einlesen menschlicher Denkmuster in Form von Gehirnscannern spielt dabei eine große Rolle. Um für den Suchmaschinenbetreiber Google entsprechende KI-Strukturen zu erhalten, kaufte man an dieser Stelle das Unternehmen DeepMind, deren Mitgründer Demis Hassabiss Neurowissenschaftler, Schachspieler und Computerspiele-Entwickler in Personalunion war. Dessen Ziel, das menschliche Gehirn in einem Computer rechnerisch abzubilden und diesem damit autonome Lernstrategien zu ermöglichen – eben jene aus Erfahrungen des Menschen – kam man bereits sehr nahe. Eine KI entstand an dieser Stelle durch klassische Atari-Konsolenspiele. Der KI wurde weder eine Anleitung noch Anweisungen zur Bedienung gegeben. Einzig und allein die Fähigkeit zu lernen wurde ihr gegeben und das Experiment gelang, die Maschine brachte sich selbstständig das Spielen von 49 Konsolenspielen bei (Tuck, 2016, S. 185).

In Anlehnung an Turing dagegen ging Kurzweil von der Genetik des Codes aus und sollte insofern recht behalten, als neueste Erkenntnisse am Gebiet der Genforschung den sog. aktiven Code der menschlichen Evolution lediglich mit 23 MB angeben (Kurzweil, 2016, S.77f). Konkret bedeutet dies, hier in 23 MB die Summe der Entwicklung des Homo Sapiens vor sich zu haben. Da heute jedes Smartphone ein Vielfaches dieses Speichers hat, müsste es hier ein einfaches

Unterfangen sein, Intelligenz zu programmieren. Dieses Unterfangen stellte sich jedoch weitaus schwieriger heraus, da andere Faktoren einfließen, v.a die Frage was Intelligenz nun konkret bedeutet. Grundlage hierfür ist das Denken. Was dieses Denken jedoch sein soll, lässt sich ebenso nicht eindeutig klären. Auch Russel und Norvig sprechen in Anlehnung an Turing davon, das vorgebrachte Argument des Bewusstseins – nämlich, wenn eine Maschine nicht nur ein Kozert komponiert, sondern auch weiß, dass sie es geschrieben hat - genauso wenig bewiesen werden kann wie die von Turing selbst formulierten Gegenfragen, ob denn Menschen denken könnten. Da auch dieser innere mentale Zustand eines Menschen bisher ebenfalls nicht hinreichend bewiesen wurde, lässt sich dies für Maschinen noch weniger beweisen (Russel, Norvig, 2012, S.1182).

Die Forschung am Menschen, wie z.B. am Genom und damit auch am Gehirn wird Kurzweil folgend noch vor Hintergrund der medizinischen Behandlungsmöglichkeiten betrachtet (2016, S.220). Auf der Hand liegen jedoch auch hier die militärischen Interessen. Hätte man damals echte Klone inklusive des Wissens von Sokrates, Newton und Einstein herstellen können, hätte man bedeutende geistige Schätze menschlichen Wissens in ihrer Gesamtheit erhalten können. Allerdings hätte man auch Duplikate von Dschingis Khan und Adolf Hitler als Kehrseite der Medaille erzeugen können. Deren Kopien hätten mit Sicherheit andere Interessen verfolgt und Menschen getötet oder töten lassen. Für Bostrom ist konsequenterweise klar, dass der Weg vom dummen Automaten bis zur Entwicklung zu einer SI und auf einen finalen Wert hin nicht bösartig oder auf zufällige Weise beschritten werden darf, sondern auf eine wohlwollend-angebrachte Art. Wie konkret dies bewerkstelligt werden soll, darauf finden sich bis heute aber keine Antworten (2014, S.276).

4.2 Menschen haben individuelle Fähigkeiten, KIs können alle haben

Denken kann als Reaktion auf diverse Aufgabenstellungen zur Bewältigung von Problemen versanden werden. Je mehr dieser Probleme von einem Menschen oder einer Maschine gelöst werden können, desto vielseitiger ist er oder sie auch einsetzbar. Folgendes Szenario sollte man sich an dieser Stelle vorstellen. Hat man Zahnschmerzen, geht man zum Zahnarzt, bei Augenproblemen zum Augenarzt. Wohin geht man aber, wenn man beide Probleme zusammen hat? Kann man hier behaupten, der Zahnarzt wäre nicht intelligent genug, Augenleiden zu kurieren und umgekehrt. Natürlich kann man! Die Lösung wäre hier eine Art Holo-Doc wie in der Science-Fiction-Serie Star Trek Voyager, der über das Wissen sowohl des Zahnarztes wie auch des Augenarztes verfügt. Auch die angesprochenen Gehirnimplantate wären denkbar, um die eigenen Fähigkeiten zu vergrößern. Somit könnte der Zahnarzt auch Augen heilen und umgekehrt. Die Idee von Kurzweil geht genau in diese Richtung, indem er davon ausgeht, dass Menschen und Maschine durch leistungsfähige Computer bedingt immer mehr zu einer Art Superintelligenz

verschmelzen. Basis für diese Idee ist Moores Gesetz, wonach sich die Rechenleistungen etwa alle zwei Jahre verdoppeln (nach Wagner, 2016, S.37).

Gefährlich in diesem Zusammenhang ist aber ein sog. Hirnschrittmacher genau dann, wenn dieser eine Persönlichkeitsänderung durch Veränderung der Stimulationsparameter herbeiführt. Dies geschah tatsächlich bei einem Parkinsonpatienten, der am helllichten Tag in der Öffentlichkeit in ein Auto einbrach. Er fühlte sich schlichtweg unbesiegbar (Clausen, 2015, S.74). Was an dieser Stelle geschehen würde, wenn eine Armee aus – noch menschlichen Soldaten – sich für unbesiegbar hält, möge sich jeder selbst beantworten. KI-technisch ist es jedoch möglich, die Stimulanzparameter per Fernsteuerung zu ändern. Seit Mitte der 90er-Jahre des vergangenen Jahrhunderts werden auch EKG-Diagramme durch Herzdiagnosen von Computern ergänzt (Kurzweil, 2016, S.123). Wenn diese KI-Prognosen falsch sind, sterben auch hier potentiell Menschen durch Falschmedikation. Hacking derartiger Systeme und dadurch bewusst falsch gelieferte Werte ist natürlich ebenfalls möglich.

Bezüglich Hackings erachtet Bostrom dies sogar als eines von sechs Kriterien für das Entstehen einer SI an, die da wären:

1. Intelligenzverstärkung, d.h. ein System kann seine Intelligenz selbst steigern.
2. Planung, im Sinne einer strategischen Planung zur Erreichung langfristiger Ziele.
3. Soziale Manipulation, um konkret Menschen zu manipulieren, die KI in die Wildbahn zu entlassen, aber auch Staaten selbst zu manipulieren - hier liegt geradezu die Gefahr, die in den Terminator-Filmen thematisiert wird begründet.
4. Hacking, im Sinne der Übernahme feindlicher KIs, der Manipulation von Finanzsystemen oder der Kontrolle der feindlichen Infrastruktur wie z.B. auch feindlichen Militärequipments.
5. Technologieentwicklung, im Sinne der Schaffung von Überwachungssystemen, schlagkräftigen Streitkräften, aber auch automatisierte Besiedelung des Weltraumes
6. wirtschaftliche Produktivität, wobei es primär darum geht, sich Einfluss, Dienstleistungen, Hardware und Ressourcen zu sichern (2014, S.134f.)

4.3 Nutznießer Militär

V.a. das US-Militär wendet Mrd.-Budgets auf, um Robotersoldaten zu entwickeln (Wagner, 2016, S.19). Laura G. Weiss vom Georgia Tech Research Institute bezweifelt ebenfalls nicht, hier Roboter schlauer werden zu lassen, allerdings wäre die Aufgabe nicht leicht, sie dann auch noch so schlau zu machen, um diese gänzlich ohne menschliche Aufsicht auszukommen zu lassen (nach Ebd., 2016, S.15f.). Was militärische Zwecke generell anbelangt, hätte es historisch betrachtet die Atom- und Wasserstoffbomben ohne Unterstützung elektronischer Rechenleistung erst gar nicht gegeben. Maßgeblich beteiligt am Manhatten-Projekt waren neben den gemeinhin bekannten Wissenschaftlern

Robert Oppenheimer und Enrico Fermi auch John von Neumann, der jedoch eher als der Vater des Computers bekannt ist, wenngleich Alan Turing die Grundlagen für diese lieferte. Erst die Berechnung von Stoßwellen durch theoretische Modelle bis zum Trinity-Test in Los Alamos machten die Bombenentwicklung möglich und führten in weiterer Folge auch 1945 zum Sieg über Japan im zweiten Weltkrieg (Schlieter, 2015, S.73). Zwar war hier noch keine KI selbst am Werke, doch die Richtung war damit vorgezeichnet, nämlich Computersysteme und damit KIs für militärische Zwecke und die Entwicklung neuer Waffensysteme zu nutzen.

Kennzeichnend für eine KI ist daher ein oben bereits erwähntes finales Endziel. An dieser Stelle wird dieses Endziel jedoch nicht im militärischen Sinne, sondern als eine zielgerichtete Handlung verstanden, von der es durchaus auch mehrere geben kann. Diese Endziele können ein Zustand sein, der eintreten soll. Uns Menschen ist es zwar nicht bewusst, aber auch wir handeln nach solchen Zielvorgaben. Wer z.B. eine numerisch bekannte Seite in einem Buch aufschlagen möchte, wird dies folgendermaßen tun: das Buch irgendwo aufschlagen, die dort vorhandene Seitennummer ansehen und danach entscheiden, ob es die gewünschte Seite ist oder doch noch vor oder nach dieser weitergesucht werden muss. Die Anzahl der möglichen Seiten wurde hier jedoch schon reduziert, da nur mehr davor oder danach gesucht werden muss. Genau dies stellt auch die Grundlage für einen Suchalgorithmus dar. Führt man sich vor Augen, dass faktisch jede Handlung eines Menschen auf die eine oder andere Art nach einem gewissen Schema abläuft, ist in Übereinstimmung mit Turings programmierbaren Maschinen sofort klar, dass jede beliebige Handlung nachprogrammiert werden kann. So gibt es heute bereits Algorithmen für faktisch alles, von obiger einfachen Suche über Navigation mittels A*-Pfadfindung (Boresch, Heinsohn, Socher, 2007, S.39) und der Vorhersage von potentiellen Kaufentscheidungen von Kunden bis hin zu Predictive Policing, um Verbrechen vorherzusagen – dazu später mehr. Die Idee hinter Algorithmen ist die Abbildung menschlichen Denkens und Handelns in allgemeingültige und allgemein anwendbare Regeln.

4.4 Die Superintelligenz hebt ab

Ausgehend von einem Basisniveau der Intelligenz des Menschen fixiert Bostrom konsequenterweise den Zeitpunkt eines Takeoffs einer KI mit dem Erreichen eines Wissensniveaus, das dem Menschen ebenbürtig ist. Ab diesem Takeoff nehmen die Fähigkeiten der KI zu bis diese eines Tages über das Wissen der gesamten Menschheit verfügt. Dann ist die Rede vom Erreichen des Niveaus einer ganzen Zivilisation. Alles was dann noch an zusätzlichem Wissen über dieses Niveau hinausgeht, ordnet man der SI zu, wobei dieses Wissen sogar aus anderen Galaxien stammen kann – daher die vorhin erwähnte Besiedelung des Weltraums bei den Kriterien der SI. Bisher liegen KIs jedoch weit unter der Schwelle des menschlichen Basisniveaus. Das eigentliche Problem dabei ist die Dauer des Takeoffs. Kennzeichnend dafür sind die Folgen

für die Gesellschaft. Erfolgt dieser langsam, kann diese reagieren und sich anpassen. Erfolgt er schnell, wobei hier durchaus nur von Minuten ausgegangen wird - im Film Terminator war es eine Nanosekunde - kann die Reaktion nicht mehr schnell genug erfolgen. Erfolgt der Takeoff dagegen gemäßigt, können Unruhen und Widerstand in einzelnen Gruppen der Gesellschaft stärker auftreten als wenn dies langsam geschieht. In allen Szenarien steht am Ende jedoch die SI (Bostrom, 2014, S.93ff) - Antworten zu deren Verhinderung sucht man an dieser Stelle jedoch vergeblich. Im Gegensatz zu Kurzweil geht Bostrom bei der SI jedoch nicht vom Verschmelzen von Mensch und Maschine aus. Genau darin besteht auch die Gefahr der Auslöschung der Menschheit. Folgende Abbildung zeigt die Entstehung der SI:

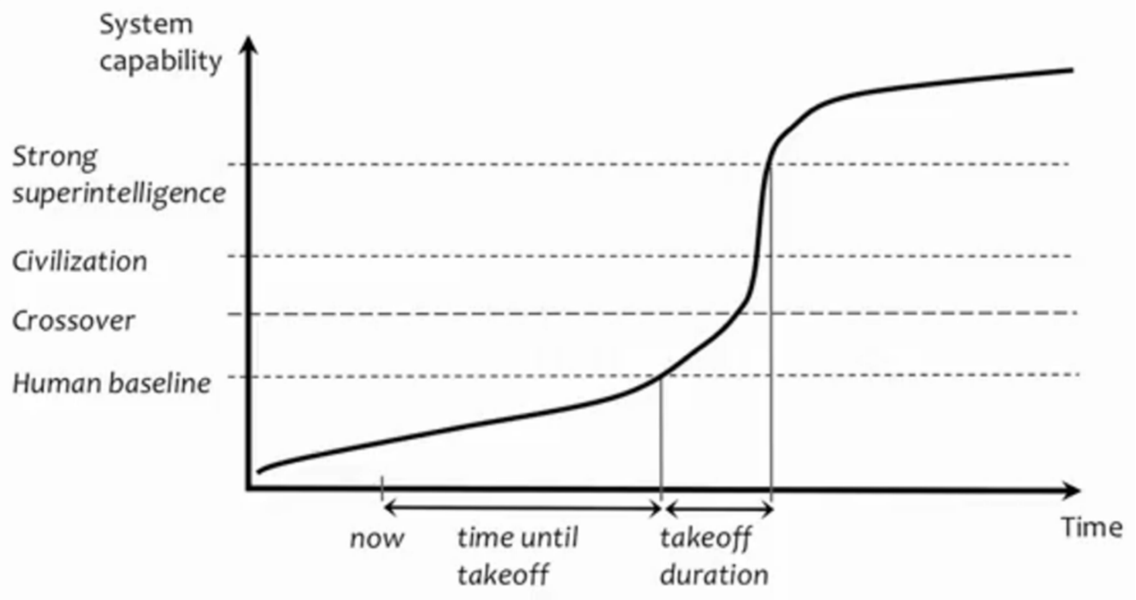


Abbildung 1: Entstehung der Superintelligenz (Bostrom [2], 2014, S.93)

4.5 KIs können missbraucht werden

Die angesprochene Gefahr hier ist allerdings nicht die KI selbst, sondern der Umstand ihres Missbrauches. Anders kann man sich die Aussagen führender Köpfe der IT-Branche nicht erklären. Ein Ende der Forschung in diesem Bereich würde hier Stillstand und Einnahmenverlust aus nicht entwickelten Technologien bedeuten. Konsequenterweise gaben etliche Experten (unter ihnen auch Hawking) in einem offenen Brief ihre Sorgen zu Protokoll. Zu den Adressaten dieser Besorgnis zählen Politik und Militär. Wenn nämlich schon Bill Gates, der frühere Chef von Microsoft Maschinen, die von Maschinen gebaut werden als düsteres Bedrohungsszenario ansieht, darf man davon ausgehen, dass die Menschheit an der Schwelle zu KI-gesteuerten Cybersoldaten steht. Besonders wenn das Turing-Prinzip nicht mehr als das Maß der Dinge bezüglich Intelligenz angesehen wird, sondern im Gegenteil die Intelligenz exponentiell ausarten wird. Dies wird weit über Turings menschliches Verhalten hinaus geschehen (Tuck, 2016, S.182ff.).

Problematisch erweist es sich zudem zu KIs keine einheitlichen Expertenmeinungen zu haben, wohin sich KIs letztlich entwickeln werden und wie lange dies dauern wird. Einigkeit besteht zwar hinsichtlich der Wahrscheinlichkeiten zur Human-Level-Machine-Intelligence (HLMI). Bis 2022 mit 10%, bis 2040 mit 50% und 2075 mit 90% (Bostrom, 2014, S.37). Ferner ist man sich einig darüber, dass KIs eine große Gefahr für die Menschheit darstellen. Neben Bill Gates sieht hier auch Tesla-Gründer Elon Musk KIs als die größte existenzielle Bedrohung der Menschheit und vergleicht sie mit Teufelsbeschwörungen. Steve Wozniak von Apple findet die Zukunft beängstigend und für die Menschen schlecht und Stephen Hawking ist der Ansicht, dass KI die großartigste, aber auch die letzte Errungenschaft der Menschen sein kann. Letztlich bringt es Paypal-Mitbegründer Peter Thiel auf den Punkt: „Eine starke KI wäre wie eine Invasion Außerirdischer. Wir würden sie nicht fragen, ob sie uns mit der Wirtschaft hilft. Wir würden sie fragen, ob sie uns umbringt." (nach Tuck, S.6).

Zusammenfassend für das Entstehen von KIs kann man an dieser Stelle sagen, diese durch Nachprogrammieren von menschlichen Denkmustern mittels einer Saat-KI zu erhalten. Dieser Saat-KI sind dabei finale Ziele mitzugeben, wobei diese dann je nachdem durch Verfolgung eines Algorithmus oder durch selbstständiges Lernen – so wie im Experiment mit den Atari-Spielen erreicht werden. In Zusammenhang mit der allumfassenden Einpflege sämtlicher Algorithmen und menschlicher Denkmuster kann es zur Intelligenzexplosion hin zu einer Superintelligenz kommen, die dann über mehr Wissen als die gesamte Menschheit zusammen verfügt und auch die Kontrolle über diese übernehmen kann. Bostrom spricht in diesem Zusammenhang auch von einer Singularität (2014, S.77). Als Gefahr für die Menschheit muss dabei eine KI nicht einmal die Werte der Menschen teilen, sondern könnte sich diesen gegenüber sogar freundlich verhalten, damit diese in ihr – vorerst - keine Gefahr erblicken. Wenn die KI einmal jedoch in freier Wildbahn schalten und walten kann, kann diese ihre wahren Absichten entfalten und die Menschen vernichten. Bostrom spircht in diesem Zusammenhang von hinterhältigem Verrat einer KI (Ebd., 2014, S.164ff.).

5 Anwendungsfälle, Möglichkeiten und Gefahren durch KIs (KIs im Alltag der IoT-Geräte)

Wer die Welt heute betrachtet, muss zwangsläufig feststellen, von KIs regelrecht umzingelt zu sein. Eine davon führen viele schon in Form eines Smartphones mit sich. Zwar handelt es sich hier um eine primitive KI wie einleitend von Hawking erwähnt, doch lässt sie sich in Kombination mit anderen KIs sehr einfach dazu benutzen, um z.B. Personen zu töten. Der Islamische Staat hat hier neu angeworbenen Kämpfern nicht zum Spaß die Smartphones abgenommen, könnten diese doch durch Kampfdrohnen geortet werden. Auch für friedliche Zwecke lassen sie sich nutzen wie gleich bei der Schwarmintelligenz zur Stauprognose gezeigt werden wird. Neben den Smartphones finden sich zudem Tablets, PCs, intelligente Kühlschränke zur selbstständigen Nachbestellung schwindender Lebensmittelvorräte, Überwachungssensoren für den Blutzuckerspiegel bei Kleinkindern, Pulsfrequenzmessgeräte, Überwachungskameras, etc... in privaten Haushalten. Diese Geräte fasst man unter dem Begriff Internet of Things zusammen und meint damit deren totale Vernetzung und deren Datenaustausch untereinander.

5.1 Killerroboter und Unsterblichkeit

Datenaustausch stellt generell das Problem bei KIs dar. Die Kombination sämtlicher gelieferten Daten von diesen Geräten zeichnet ein entsprechendes Persönlichkeitsprofil betreffender Personen. Hier geht es jedoch nicht um Data-Mining, sondern um die sich abzeichnenden Gefahrenmöglichkeiten der Übernahme dieser KI-Systeme durch andere Maschinen, denn was die Möglichkeiten von KIs dabei anbelangt sind diese schier unbegrenzt. Was die Gefahren allerdings anbelangt, sind sie es jedoch auch.

Im letzten Kapitel wurden bereits Aussagen führender Köpfe der IT und Wissenschaft angesprochen. Sämtlichen dieser Aussagen war zu entnehmen, sich mit KIs der Gefahr auszusetzen, den Maschinen eines Tages die Machtübernahme zu ermöglichen. Ängste herrschen hierbei bei Bill Gates zusätzlich noch vor, indem dieser nicht versteht, warum nicht noch mehr Menschen beunruhigt sind. An dieser Stelle muss die Frage aufgeworfen werden, warum denn hier führende IT-Experten derart düstere Szenarien aufzeigen, sehr wohl allerdings ihre Produkte in diese Richtung entwickeln. Warum stellt Microsoft Software her, warum baut Tesla Autos, warum Apple Computer und warum wird Paypal als Bezahldienst genutzt? Die Menschen sind doch verrückt, wenn sie sich von diesen Systemen bevormunden oder gar töten lassen. Eine mögliche Erklärung wäre, hier durch KIs einen Vorteil zu ziehen. Damit sind natürlich wirtschaftliche Beweggründe gemeint. Die Technologie lässt sich jedoch einfach missbräuchlich verwenden und an Waffensysteme adaptieren. Wagner führt hier treffend Roboterwaffen und Selbstschussanlagen an, aber auch Minen. Bei diesen Systemen gibt kein Mensch den Befehl zum Abschuss, sondern das System selbst löst sich anhand

vorgegebener Umstände aus. Mehr noch, sieht er eben diejenige Partei im Vorteil, die sich zuerst der extrem hohen Verarbeitungsgeschwindigkeit von KIs bedient: „Der erste, der sich der Killerroboter bedient, gewinnt also einen entscheidenden Vorteil“ (2015, S. 113f.) Anzumerken wäre, hier die Steuerung dieser Killerroboter eben jenen Systemen zu überlassen, wie sie auch am Consumer-Markt angeboten werden, d.h. konkret also letztlich der Hard- und Software obiger Hersteller. Auch Kurzweil denkt hier ähnlich und sieht die „fortschrittlichen Nationen“ dabei im Vorteil (2016, S.274), nennt diese jedoch nicht explizit beim Namen. Es liegt allerdings auf der Hand, dass hier v.a. die USA gemeint sind.

Wenn das Szenario, was Kurzweil weiter vorschwebt, jedoch Realität wird, existiert Sterben im klassischen Sinn aber ohnehin nicht mehr. Unsterblichkeit lässt sich – so glaubt zumindest Kurzweil – wie im Film Transcendence in einigen Jahren tatsächlich realisieren. Ob ihn dabei eine Star Trek-Folge inspiriert hat, wo das Wissen eines Verstorbenen in den Androiden Data übertragen wurde oder das zylonische Auferstehungszentrum aus Battlestar Galactica, wo von einem gefallenen Zylonen sofort der Gedächtnisdownload in einen Klon erfolgt, ist leider nicht bekannt. Das Ziel jedenfalls ist genau ein solcher Upload des Intellekts in einen Supercomputer spätestens im Jahr 2029 (nach Tuck, 2016, S.214f.). Ob dies das ewige Leben ist, sei dahingestellt. Kurzweil selbst führt hierzu an:

„Wenn wir Software sind, wird unsere Existenz nicht mehr von der Lebensdauer unserer Daten verarbeitenden Schaltungen abhängen. Es wird immer noch Hardware und Körper geben, aber im Kern wird unsere Identität in der Permanenz unserer Software liegen. Genau wie wir heute unsere Dateien nicht wegwerfen, wenn wir uns einen neuen Personal Computer kaufen, sondern alles, was wir behalten wollen, auf das neue Gerät übertragen, so werden wir auch unsere Bewusstseinsdateien nicht wegwerfen, wenn wir uns in regelmäßigen Abständen auf den jeweils neuesten, jedes Mal leistungsfähigeren "persönlichen" Computer übertragen“ (2016, S.211).

Was Kurzweil an dieser Stelle jedoch nicht bedenkt, ist der Untergang des ursprünglichen Körpers samt seiner Seele. An dieser Stelle werden moralische, ethische oder religiöse Fragen aufgeworfen. Sie zu beantworten wird noch Generationen von Forschern und Priestern beschäftigen. Raoúl Rojas, Informatikprofessor der FU Berlin ist hier der Ansicht, dass dabei Robotik dem Computer lediglich den entsprechenden Körper gibt. Der Zweck ist, dem Computer damit das Handeln und auch weiteres Lernen zu ermöglichen. Ray Kurzweil hingegen hält den Nachbau eines menschlichen Gehirns selbst derzeit zwar noch nicht für möglich, lässt auf seiner eigenen Singularity Universität in Kalifornien aber intensiv an KIs forschen, die Komplexität des menschlichen Gehirns zu ergründen (nach Wagner, 2016, S.61f.). Abwehrstrategien werden am MIRI entwickelt. Kritik gibt es dabei an der Singularity Universität. Während man dort nämlich an deskriptiven Projekten zur Superintelligenz arbeitet, bemüht man sich am MIRI dieser, wenn es soweit ist, rechtzeitig

humane Werte zu vermitteln. Anderenfalls drohe die Auslöschung der Menschheit, vor der bereits Hawking gewarnt hat (Ebd., 2016, S.80f). Welche Fragen hinsichtlich des religiösen Glaubens, der Moral und Ethik damit verbunden sind, muss erst die Zukunft zeigen. Man stelle sich aber an dieser Stelle vor man begegnet sich selbst und dieses zweite Ich tötet einen mit der Begründung, wonach man selbst nur eine Kopie sei, obwohl es eigentlich die Kopie ist, die sich für das Original hält. Deutlich wird jedenfalls der interdisziplinäre Zusammenhang in der KI-Forschung.

5.2 Natur als Vorbild

Hinsichtlich der Forschung muss bei den bereits erwähnten KI-Agenten zwischen der Bewegung und der eigentlichen Handlung unterschieden werden. Teahan geht hierbei von grundsätzlichen Unterschieden in den Lebewesen aus und weist treffend auf unterschiedliches Verhalten in unterschiedlichen Staaten, Gruppen oder Personen hin. Auch verhalten sich Tiere anders als Menschen. Ferner unterscheidet Teahan zwischen psychologisch bewusst gesetzter Handlung oder solchen Handlungen, die durch Instinkte erfolgen (2010, S.10). Bezogen auf KIs ist dies dann wichtig, wenn jegliche Art von Verhalten eines Wesens durch eine KI nachgebildet werden soll. Im militärischen Bereich wäre hier z.B. an künstliche Vögel oder miniaturisierte Insekten zu Spionagezwecken zu denken. Ein derartiger Vogel, der sich nicht wie ein natürlicher Vogel verhalten würde, würde sofort die Aufmerksamkeit erregen. Auch ein starr vor sich herschwebendes künstliches Insekt würde auffallen.

Von besonderer Tragweite für Verhaltenssteuerung, egal ob künstlicher Vogel, Insekt oder militärische Drohne sind dabei Algorithmen zur Zielfindung. In der Literatur finden sich diesbezüglich die bekannten Algorithmen zur Pfadsuche, wie der bereits erwähnte A*-Algorithmus. Je nachdem bereits informiert über die jeweiligen Kosten, im Sinne von der Distanz zum Ziel oder nicht käme auch Dijkstra infrage (Teahan, 2010, S.127ff.). Technisch laufen diese Algorithmen über verschiedene Knotenpunkte, die expandiert und dahingehend ausgewertet werden, wie weit man noch vom Ziel entfernt ist und ob weitere Wege von den Nachfolgeknoten zum Ziel existieren. Ist dem nicht so, werden bisher noch nicht berücksichtigte Knoten einbezogen. Weitaus interessanter als die eigentliche programmtechnische Umsetzung ist jedoch, derartige Algorithmen bereits in faktisch allen Computerspielen vorzufinden sowie auch in Saugrobotern. Wenn also von der erwähnten Saat-KI ausgegangen wird, die sich einmal zu einer Superintelligenz entwickelt, dann handelt es sich bei derartigen Algorithmen eindeutig um eine solche Saat. Doch schon in der Bibel heißt es, wer den Wind sät, wird den Sturm ernten. Nun stelle man sich vor, sämtliche bekannten Algorithmen der Welt wären in dieser SI vereinigt. Sie könnte sowohl Aussagen zu Flugbahnberechnungen treffen, womit an das Abfangen einerseits und optimales Steuern von Raketen andererseits zu denken wäre. Sie könnte aber auch Stimmen imitieren. Akustische Signale und Sprache

sind ja bekanntlich physikalisch nichts anderes als Frequenzwellen mit einer bestimmten Amplitude, Länge und Intensität. Vor diesem Hintergrund betrachtet bekommen die einleitend erwähnten Terminater-Filme konkrete Gestalt. Die Szenarien, die von Elon Musk über Bill Gates bis Stephen Hawking hinsichtlich der Gefahren von KIs somit aufgezeigt wurden, sind daher als real einzustufen. Und mehr noch, denn die Wiederkehr des Heilands wird nicht in Fleisch und Blut, sondern in Form von Silizium, Plastik und Metall erfolgen (nach Wagner, 2016, S.40f.)

5.3 Internet of Things „Anwendungen“

Weitere Anwendung finden KIs neben dem militärischen Bereich auch eine Stufe darunter bei der Polizei in der Vorhersage von Verbrechen, noch bevor diese Geschehen. Wer hier denkt, auch dies wäre Science-Fiction aus dem Film Minority Report, der sei eines Besseren belehrt. Es ist Realität. Überwachung durch vernetzte Geräte und deren Sensoren lässt sich hervorragend genau dazu nutzen. Das Ganze läuft in den USA unter dem Begriff des „Predictive Policing“ (Bambusch, 2015, S.25). Bedenkt man diesbezüglich, dass militärische Drohnen auch im Inland der USA gegen die Rauschgift-Mafia und illegale Einwanderer von Mexico eingesetzt werden, ist es nur eine Frage der Zeit, bis hier - rein irrtümlich natürlich - eine Person getötet wird, möglicherweise auch eine, die gar nicht ins Zielraster passt (Bambusch [2], 2012, S.102).

Alex Pentland vom MIT spricht hier geradezu von einer neuen Wissensdisziplin, nämlich von Sozialphysik. Mittels Sensoren der vernetzten Geräte wird konsequenterweise das menschliche Sozialsystem überwacht. Dies geht sogar so weit, als sich mittels an Tatorten zurückgebliebener DNA-Spuren ganze Phantombilder betreffender Verdächtiger generieren lassen (nach Schlieter, 2015, S.41f). Anzumerken ist allerdings, hier sehr schnell falschen Verdächtigungen ausgesetzt sein zu können, wenn KIs Bilder erstellen, die wie Allerweltsbilder aussehen. Auch Smartphones wurden bereits als Systeme herangezogen, die vor Gericht faktisch „gegen einen aussagen“, besonders wenn dort Informationen gespeichert sind, die belastendes Material für Straftaten beinhalten (Tuck, 2016, S.179).

Gefahren birgen auch Schreibprogramme in sich. Hier ist aber nicht die Autozusammenfassen-Funktion der Textverarbeitung Word gemeint, sondern das Schreiben von Nachrichten durch KIs. Unternehmen nutzen dies auch für Börsenberichte. Was passiert, wenn falsche Börsenkurse angegeben werden oder um es mit den Worten von US-Präsident Trump auszudrücken generell „Fake News“ veröffentlicht werden, kann man sich spätestens dann vorstellen, wenn naturwissenschaftliche, technische oder medizinische Artikel in Fachzeitschriften automatisch generiert werden. Besonders bei medizinischen Artikeln sind einmal mehr Menschenleben gefährdet, wenn das dort publizierte Wissen eigentlich Unsinn darstellen würde. Denkbar wären aber auch Angriffe auf Server von Verlagshäusern und bewusste Änderung von Informationen

(Mainzer, 2016, S.78). An dieser Stelle wird deutlich, dass faktisch sämtliche Lebensbereiche von KIs durchzogen sind. Wie einleitend erwähnt wurde, handelt es sich bei der Mehrzahl zwar um die von Hawking angeführten primitiven KIs. Diese einzelnen KIs für sich stellen dabei noch keine Gefahr dar, deren Kombination allerdings sehr wohl, v.a. dann, wenn Bostroms angesprochene SI diese eines Tages „übernimmt".

5.4 Schwarmintelligenz zu militärischen Zwecken

Mit einer Kombination einzelner, kleinerer KIs lässt sich deren Leistung im Zuge von Schwarmintelligenz steigern. Damit ist die Möglichkeit gemeint, mittels individuellen, lokalen Intelligenzen auf ein gemeinsames großes Ziel hinzuarbeiten. Entscheidend ist dabei die Koordination. Aus der Natur lassen sich hier z.B. Bienenvölker und Ameisenkolonien treffend anführen. Während bei Ameisen die Masse mittels ihrer Duftmarken anderen den kürzesten Weg zum Ziel markiert, weisen Bienen mit entsprechenden Tanzmustern darauf hin, wo Blütenpollen und damit Honig wartet. Anhand der immer intensiver werdenden Ameisenduftmarken kumuliert sich bald eine viel belaufene Ameisenstrecke, während durch immer mehr Bienentänze auch immer mehr andere Bienen auf die Wiesen aufmerksam werden. Nutzbar macht man sich dieses instinkthafte Schwarmverhalten, also mit der Masse mitschwimmen, z.B. bei eingebuchten Handys zur Stauprognose. Je mehr eingebuchte Handys in einer Funkzelle still stehen - Handys lassen sich nämlich einfach orten - desto höher ist die Wahrscheinlichkeit, hier einen Stau vor sich zu haben. Diese Kumulation von Daten hat hier nichts mit Data-Mining zu tun, sondern einfach ausgedrückt nur mit Nutzung von Schwarmintelligenz zur Leistungssteigerung der Masse. Diese profitiert dann von der kollektiven Intelligenz kleinerer Systeme: nämlich den Schluss zu ziehen, wo es sich staut und diesen umfahren zu können (Münkel, 2012, S.114ff.). Künftig kann diese Schwarmintelligenz auch gleich ein autonom agierendes Fahrzeug dazu nutzen.

Schwarmintelligenz ist prinzipiell geradezu dann nützlich, wenn Ausfallsicherheit gegeben sein soll. Denn nicht die Ameisen sind es, die z.B. bei Amazon die Lagerorganisation abwickeln, sondern Roboter. Entwickelt wurde das dortige System durch ein Unternehmen der TU München. Dieses kann auch in klassischen Lagern zur Anwendung kommen, wobei die eingesetzten Roboter mit Namen Toru ebenfalls im Zuge von Schwarmintelligenz organisiert sind. Wie Eberl ausführt, kann dabei jederzeit ein anderer Roboter die Aufgaben einer ausgefallenen Einheit übernehmen (2016, S.203ff.). Damit ist ein klassisches redundantes Backup-Konzept gemeint. Für Personen gefährlich ist dies allerdings, wenn sie es mit einer Masse an autonomen Roboter-Systemen zu tun bekäme, bei denen salopp ausgedrückt „eine Schraube locker" wäre. Eine allenfalls ausgeschaltete Einheit würde sofort durch eine andere ersetzt werden. Fehlen hier Notfall-Stopp-Mechanismen, wie im Kapitel zur Fehlervermeidung gezeigt wird, liegt auf der Hand, es hier nicht mit abwegigen

Szenarien zu tun zu bekommen, in denen sich Roboter gegen den Menschen wenden könnten.

Sind diese Roboter so klein, dass man sie nicht erkennen kann, liegt eine ganz neue Dimension von Waffe vor. Von diesen Mikrowaffen gibt es an der Harvard-Universität in Boston bereits Prototypen. Damit wird deutlich welches Näheverhältnis das US-Militär zur Forschung hat. Auch hier ist Schwarmintelligenz erforderlich, jedoch in Zusammenhang mit dem gemeinsamen, koordinierten Zusammenfliegen und einander ausweichen. Tuck spricht in diesem Zusammenhang treffend von Spionagezwecken derartiger Mikrowaffen bis hin zu Sabotageakten an feindlicher Infrastruktur, wie z.B. Serverräumen in Rechenzentren. Werden nämlich feindliche KIs durch Mikrowaffen ausgeschaltet, wäre hier der Weg frei für Militärschläge ohne Gegenwehr. Für Tuck könnten dabei die Mikrowaffen - da sie derzeit noch über kurze Reichweiten verfügen künftig anstatt Bomben aus Stealth Bombern über feindlichem Gebiet abgeworfen werden (Tuck, 2016, S.87). Auch hier merkt man Parallelen zu den Terminator-Filmen, wo darauf hingewiesen wird, wie sämtliche Stealth Bomber völlig unbemannt fliegen werden. Tuck weist noch auf die Effizienz der Maschinen hin, da diese robuster als Menschen sind und daher gewagtere G-Manöver fliegen können, zudem nicht Stimmungsschwankungen unterliegen und weder Schleudersitze noch Atemluft in großen Höhen benötigen. Zudem müssen sie nie pausieren (Ebd., 2016, S.83f). In Zusammenhang mit autonom fliegenden Stealth Bombern wäre dies die vollständig autonome Kriegsführung. Gefahr für Menschenleben besteht dabei besonders dann, wenn einzelne Mikrowaffen außer Kontrolle geraten sollten. Angesichts ihres Auftretens in Schwärmen ist dieses Szenario nicht abwegig. Man denke hier nur an Hunderte von KI-gesteuerten Mikroinsekten, ausgestattet mit Minisprengsätzen. Völkerrechtlich geächteten Streubomben stehen diese nämlich um nichts nach. Denkt man sich hier noch autonome Tankflugzeuge hinzu, könnten permanent KI-gesteuerte Überwachungs- und Angriffsbomber in der Luft umherkreisen.

In die Ähnliche Richtung weist Kurzweil hinsichtlich Nanobots, indem er auf deren Gefahr hinweist, wenn Millionen von kleinen KIs aufeinanderprallen. Hier droht ein Szenario, das einfach nicht mehr kontrollierbar ist (2016, S.229). Im Bereich der Nanobot-Schwärme schweift Kurzweil jedoch aus heutiger Sicht noch in den Bereich der Science Fiction ab - offenbar denkt er an die TV-Serie Stargate und die dortigen Replikatoren, wenn er darauf hinweist, dass Nanobots augenblicklich jede gewünschte Form und Persönlichkeit annehmen können (2016, S.241). Angesichts der militärischen Bedeutung von KIs verwundert es aber nicht, mittels z.B. getarnter Steine, in denen in Wahrheit Sensoren stecken, das Leben zu überwachen und allfällige Angriffsziele auszumachen. Tuck spricht in diesem Zusammenhang von ziviltechnischem Nutzen für den Verkehr, um hier auf Druckänderungen zu reagieren, bzw. die Steine als Bewegungsmelder einzusetzen. Militärisch dagegen können diese

Sensoren gepaart mit intelligenter Software Datenmassen organisieren und analysieren. Anzumerken ist hierzu, natürlich die militärische Überwachung im Sinn zu haben. Wie Tuck jedoch weiter einräumt kann bei derartigen Sensoren auch viel schiefgehen - womit dann natürlich unschuldige Menschenleben gefährdet werden:

„In der Praxis kann bei ferngesteuerten Sensoren viel schiefgehen. Elektronik kann versagen. Mikrofone können verschmutzen Kameras können mit der Linse nach unten landen. Was dann passiert, ist Aufgabe der Software." (Tuck, 2016, S.92)

Damit wurde auf den Punkt gebracht, wodurch KIs versagen können, nämlich durch Hardware- und Softwareprobleme. Das Risiko, Opfer einer fehlerhaft arbeitenden KI zu werden steigt entsprechend, je mehr Hardware in einer Komponente verbaut und je mehr Codezeilen in der Steuersoftware vorhanden sind. Besonders im Zeitalter des Internet of Things stellt eine Kontrolllogik an faktisch jeder Ecke eine Gefahr dar, hier potentiell zu versagen und dadurch Menschenleben zu gefährden. Man sollte sich an dieser Stelle fragen, was passieren würde, wenn Verkehrsleitsysteme abstürzen, die Stromversorgung zusammenbricht, Flugzeugnavigation ausfällt oder Waffensysteme eine Eigendynamik entwickeln.

Teahan merkt treffend zur Schwarmintelligenz von Ameisen an, wie deren Verhalten bezüglich Problemlösungsstrategien auf Computer zu übertragen wäre, nämlich mit verteilten Datenstrukturen und –speichern. Diese können dann von anderen abgerufen werden, also ganz genau wie bei der Schwarmintelligenz das einzelne Individuum vom kollektiven Wissen des Ganzen profitiert.

„Stigmergy is not restricted to natural life examples - the Internet is one obvious example. Many computer systems also make use of stigmergy - for example, the ant colony optimization algorithm is a method for finding optimal paths as solutions to problems. Some computer systems use shared data structures that are managed by a distrubuted community of clients that supports emergent organization. One example is the blackboard architecture as used in AI systems first developed in the 1980s. A blackboard makes use of communication via a shared memory that can be written to independently by an agent then examined by other agents much like a real-life blackboard can in a lecture room. Blackboards are now being used in first-person shooter video games, and as a means of communication between agents in a computer network." (Teahan, 2010, S.23)

Damit hat Teahan exakt das beschrieben, was Bostrom meint, nämlich den Takeoff und die Expansion sämtlichen Wissens in eine SI. Problematisch wird es allerdings dann, wenn die in Teahans Zitat erwähnten Agents die Kontrolle übernehmen. Es stellt sich somit die Frage, warum denn im Zeitalter des Internet of Things, immer mehr Geräte mit KI ausgestattet werden, wenn die Gefahr der Übernahme der Kontrolle durch die Maschinen in immer höherem Maße gegeben ist, je mehr Geräte hier den potentiellen Schwarm bilden. Antworten auf diese Frage wären hier jedoch rein spekulativer Natur.

Thematisiert wird die Problematik allerdings im Animationsfilm „Wall E“ von 2008, wo die Menschheit durch ihre Bequemlichkeit ebenfalls in die Abhängigkeit ihrer Maschinen gerät. Maschinen haben die Kontrolle über alle Lebensbereiche mit der Wirkung, jedwedes Schöpferische im Menschen ausgeschaltet zu haben. Stellt man sich die Frage, was Maschinen heute kontrollieren, so lautet die Schlussfolgerung: einfach alles und dabei auch kritische Bereiche – wie eben von der Elektrizitätsversorgung, über Telekommunikationsnetze, Steuerung der Wasserwerke, Treibstoffpumpen, Autopiloten und Verkehrsnetze.

5.5 Autonomes Fahren

Was die Verkehrsnetze anbelangt und die Anwendung auf der Straße in autonomen Fahrzeugen, erfordert dies konsequenterweise auch zuverlässiges Erkennen von Verkehrszeichen. Diesbezüglich gewann Prof. Schmidhuber im Jahre 2011 mittels Deep Learning beim richtigen Erkennen von 50.000 Verkehrszeichen den Wettbewerb für Neuroinformatik in Bochum. Schmidhubers KI schnitt mit einer Fehlerrate von 0,54% besser ab als eine Kontrollgruppe von 32 Menschen, die 1,16% Fehlerrate aufwies (Eberl, 2016, S.107). Außer Acht lassen darf man an dieser Stelle jedoch nicht den menschlichen Faktor in echten Verkehrssituationen. Würde nämlich das Straßenverkehrszeichen Vorrangstraße korrekt erkannt werden und sich von einer Seitenstraße ein Fahrzeug mit hoher Geschwindigkeit nähern, liegt es in der Natur des Menschen hier präventiv abzubremsen, da das Fahrzeug ja den Vorrang missachten könnte. Eine KI dagegen würde hier strikt nach Vorschrift vorgehen und auf Vorrang setzen. Einwenden könnte man an dieser Stelle natürlich, der KI auch weitere Sensoren hinzuzufügen und ein von der Seite sich näherndes schnell fahrendes Fahrzeug als Kollisionsbedrohung anzusehen. Hierbei ergeben sich aber erhebliche Schwierigkeiten hinsichtlich der Mustererkennung. Was würde nämlich passieren, wenn das Fahrzeug hinter Büschen fahren würde, wo ein Mensch eindeutig zwar noch ein Fahrzeug erkennen könnte, eine KI dagegen jedoch noch überfordert wäre. Das Problem auf den Punkt bringt Eberl, indem er auch hier den menschlichen Verstand anführt. Weht hier z.B. der Wind eine Papiertüte gegen die Windschutzscheibe, würde kein Mensch auf den Gedanken kommen, eine Vollbremsung einzuleiten und möglicherweise Auffahrunfälle nachfolgender Fahrzeuge zu riskieren. Einer KI jedoch den Unterschied beizubringen zwischen einer gefahrlosen leichten Papiertüte, mit der die Kollision durchgeführt werden darf, mit einem schweren Objekt jedoch nicht, stößt an die Grenzen des Machbaren (Eberl, 2016, S.124).

Ein ungelöstes Problem ist es zudem das Erlernen von moralischen Werten. Wenn z.B. vor einem autonom fahrenden Fahrzeug plötzlich ein Hindernis in Form einer Mutter und einem Kind auftaucht, die KI-Steuerung des Fahrzeuges jedoch ermittelt, nicht mehr rechtzeitig anhalten zu können, was soll dann geschehen? Soll das Kind überfahren werden, die Mutter oder soll sich das

Fahrzeug selbst vernichten samt dem Lenker, indem es ausweicht und über ein Kliff stürzt (Alpaydin, 2016, S.165)? Die Beantwortung dieser Frage ist bisher noch ungelöst.

5.6 BOT-Netze als Einfallstor für die KI-Übernahme

Da sämtliche heutige KI-Geräte auch schon einmal gehackt wurden, erscheint es nur logisch, wenn Antiviren-Software-Hersteller wie McAfee Gründer John McAfee ebenfalls vor KIs warnt indem er davon ausgeht, Alltagsgegenstände wie Kühlschränke, Safes und Thermostats als weitaus gefährlicher einzustufen als Smartphones oder Computer. Letztlich ist deren Anzahl ja viel zahlreicher. V.a. erachtet McAfee China als Feind, da viele Alltagsgeräte dort gefertigt werden und über das Internet im Zuge von Cyberkriegen dann als BOT-Netze gegen andere Staaten eingesetzt werden könnten. (ORF, 2016). Anzumerken ist, wenngleich hier durch die von Menschen ausgelösten Cyberangriffe Kontrollübernahmen erfolgen, über die Schaltungen in diesen Geräten, sehr wohl Menschen geschädigt und getötet werden können. Dazu reicht es aus, einfach ein Relais per Softwaresteuerung zu schalten um die Außenhülle eines IoT-Gerätes unter Spannung zu setzen und damit einer Person einen tödlichen Stromschlag zuzuführen. Bezogen auf den einleitenden Kaffeeautomaten könnte auch dieser anstatt Kaffee, Giftpulver verwenden. Die Gefahr besteht somit im grundsätzlichen Schaffen von Infrastruktur durch den Menschen, die zwar noch unter staatlicher Kontrolle steht in Erwartung von Cyberkriegen gegen andere Staaten. Wenn diese jedoch eines Tages unter der Kontrolle einer KI, möglicherweise einer SI steht, ist genau das Szenario der Terminator-Filme eingetreten. Die Gefahr geht dabei von sog. BOT-Netzen aus. Derzeit werden diese noch durch Menschen gesteuert und übernommen, um bei Netzwerken dahingehend Schäden zu verursachen, indem diese lahmgelegt werden, was mittels massenhafter Anfragen sämtlicher übernommenen Geräte an das betreffende Netzwerk, bzw. die betreffenden Server geschieht. Das Ergebnis ist hier der Zusammenbruch unter der Last, also ein Denial of Service (Hosbach [2], 2015, S.46ff).

6 Wie lernen Maschinen?

Zur Beantwortung wie Maschinen lernen, ist zuerst das Verständnis wie Menschen lernen, zu klären. Menschen lernen hier klassisch von den Eltern einen zentralen Wortschatz (Sprache) durch Zuhören als Baby. Als Kleinkind eignet sich der Mensch bereits rudimentäres Weltwissen an, z.B. wie Bilder in einem Buch auch dann noch erhalten bleiben, wenn es geschlossen wird. Die Einteilung von Objekten in Klassen sowie Abstraktion rundet die Bildung ab. Treffend wirft Eberl hier die Frage auf, ob denn nicht auch KI-gesteuerte Roboter zusammen mit Menschen „aufwachsen" müssen, um Erfahrungen zu sammeln (2016, S.63ff.).

Im Kindergarten wird sozialer Kontakt erlernt, in der Volksschule grundlegende Schriftzeichen (Alphabet) und Syntax (Grammatik), später folgen vertiefende Details in den Gymnasien, bzw. Spezialwissen an Universitäten oder Fachhochschulen. Hier spielt somit der menschliche Verstand die entscheidende Rolle. Was Verstand ist, ist jedoch für Maschinen nicht ersichtlich. Hier verhält es sich wie mit einem Kleinkind, das sowohl einen Vogel wie ein Flugzeug erblickt. Für das Kind existiert beides seit jeher und fliegt. Dass der Vogel dies von Natur aus tut, das Flugzeug allerdings erst nach der Entdeckung der Gesetze der Aerodynamik und der Entwicklung durch den Menschen, kann das Kind erst im Laufe der Zeit erlernen. Diesen Unterschied allerdings einer Maschine einzupflanzen wäre der Verstand. Und einen solchen gibt es dafür noch nicht (Alpaydin, 2016, S.22).

Abgestimmt wird hier auf die sensomotorische Lernfähigkeit. Diese ist allerdings eingeschränkt, da es derzeit noch nicht vorstellbar ist etwa 900 Mio. tastsensible Rezeptoren des Menschen in einer KI zu implementieren. Selbst der besten KI würden durch fehlende Sensoren einfach Informationen verloren gehen, die der Mensch hier erhalten würde. Ein Objekt durch mangelhafte oder weniger Sensoren zu analysieren würde allerdings einer KI kein richtiges Abbild davon vermitteln. Im Unterschied zur KI steht zudem auch die menschliche Neugier im Fokus der Betrachtung. Während eine KI auf Input wartet, wartet der Mensch nicht, sondern greift zu, um etwas zu analysieren, so Roboterforscher Rolf Pfeifer. (nach Eberl, 2016, S.79).

Entscheidend im Lernprozess für KIs ist es daher, den Verstand in diese auf die eine oder andere Weise zu implementieren. Beste Überwachungstechnik nützt nämlich nichts, wenn sie, wie Fei-Fei Li, Expertin für computergestütztes Sehen und maschinelles Lernen an der Stanford University einräumt, einen ertrinkenden Menschen im Schwimmbad nicht melden könnte. Hier muss also weiterhin der menschliche Bademeister achtsam bleiben. Der Umgang mit extern erhaltenen Daten und deren semantische Interpretation stellt in diesem Zusammenhang Georg Gottlob, Informatik-Professor in Oxford zufolge, geradezu den heiligen Gral der KI dar. Daten müsste hier eine Bedeutung zugeordnet werden, wobei genau dies die eigentliche Schwierigkeit im binären

Denken darstellt. Wie soll der Duft einer Rose einer Maschine beigebracht werden? Sie könnte diesen zwar in alle möglichen chemischen Verbindungen aufschlüsseln und analysieren, entsprechende Sensoren vorausgesetzt, damit jedoch Gefühle zu verbinden, liegt außerhalb deren „Verstandes“ (nach Eberl, 2016, S. 124f).

Weitere konkrete Lernmethoden stellen darauf ab, Ähnlichkeiten in Mustern herauszufinden. Alpaydin bringt hier Beispiele des Gebrauchtwagenmarktes, wobei eine Maschine lernen kann, bei ähnlichen Baujahren und Kilometerständen ähnliche Preise zu erzielen. Gesucht ist also ein allgemeingültiges Modell zur entsprechenden Vorgehensweise. Ohne o.a. Verstand ist dieses Unterfangen jedoch sehr schwierig. Davon zeugen die Attribute eines Fahrzeuges, wie ob es garagengepflegt war, stark verrostet oder ein Unfallwagen ist (Alpaydin, 2016, S.34ff.).

6.1 Schach-KIs und Entscheidungsbäume

Wichtig für Entscheidungsfindungen sind im Bereich der KI, genau wie in der Informatik schlechthin, Entscheidungsbäume im Sinne von Wenn/Dann-Strukturen. Ausgehend von einem Wurzelelement werden immer weitere Verzweigungen durchsucht, bis letztlich ein Problemlösungsweg gefunden wurde. Problematisch daran erscheint, dass man zuerst einmal einen solchen Entscheidungsbaum generieren muss, was wiederum Lebenserfahrung und deren Abbildung auf eine KI bedingt. Auch die Suche nach vorerst unbekannten Entscheidungsalternativen wäre hier zu nennen. In beiden Fällen wird deutlich, wie lückenhaft Entscheidungsbäume sein können, wenn man eine entsprechende Erfahrung noch nicht gemacht hat oder auf der Suche danach bisher noch nicht vorhandene Ereignisse im späteren Verlauf eintreten (Alpaydin, 2016, S.77f). Verfügt man einmal über einen theoretisch hinreichend tiefen Entscheidungsbaum mit dem Gesamtwissen des Universums, ist es keine Kunst mehr Antworten auf jede Art von Problem zu erhalten. Wie Segaran anmerkt, erhält man Antworten ausgehend vom Wurzelelement, indem man immer weiter nach unten entlangfährt. Begründungen dagegen erhält man, indem man die entgegengesetzte Richtung einschlägt (2008, S. 161). Damit ist dann entweder ein Top-down- oder Bottom-up-Ansatz gemeint. Kennzeichnend für solche Entscheidungsbäume und Lösungsstrategien ist dabei geradezu das Schachspiel. An dieser Stelle können zwar nicht die Schachregeln erläutert werden, diese sind allerdings für das Problemverständnis bezüglich KIs nicht wichtig. Es geht allerdings darum, den gegnerischen König mittels der eigenen Figuren matt zu setzen, d.h. ihn mit einer Figur zu bedrohen und gleichzeitig alle Fluchtmöglichkeiten auf ein anderes Feld zu nehmen.

Wie Steinwender und Friedel treffend anführen, existiert hier keine Möglichkeit, einem Computer dabei den Unterschied zwischen guten und schlechten Zügen beizubringen. Zudem gibt es keine vollständige Symmetrie

des Schachbrettes, da König und Dame in der Grundstellung in der gleichen Reihe stehen. Schach ist mit 64 Feldern und 32 Figuren um ein Vielfaches komplexer als das gleich besprochene TTT. So existieren bei 80 Halbzügen 1,5x10 hoch 128 mögliche Spielkombinationen, also 3,464823841570940447511989197632 6e+150 Stellungen. Führt man sich vor Augen, dass die Anzahl der Elementarteilchen im Universum auf 3,2x10 hoch 78 geschätzt wird, liegt deren Anzahl noch deutlich darunter: 2,52172839656924666958585664092e+117. Interessant in diesem Zusammenhang, jedoch ungesichert erscheint der erzwungene Gewinn durch Weiß. So soll bereits 1975 unter Leitung von KI-Forscher Richard Pinkleaf am MIT ein Schachcomputer entwickelt worden sein, der pro Sekunde 1,5 Partien gegen sich selbst spielte. Nach sieben Monaten Dauerbetrieb zeichnete sich der Eröffnungszug h2-h4 bei optimaler Spielstrategie als Gewinnzug für Weiß ab. Der damalige Weltmeister Bobby Fischer gab allerdings an, diese Eröffnung widerlegt zu haben. (Steinwender, Friedel, 1995, S.50ff).

Warum es hier weniger schwer ist, eigentliche Algorithmen für die Bewegung der Schachfiguren und die Spielregeln zu schreiben, denn dahinter steckende Strategien zu entwickeln, liegt im erforderlichen Speicher begründet. Wie o.a wurde, existieren mehr als astronomisch hohe Variationen von Schachstellungen. Die Kunst ist es daher, einen Entscheidungsbaum zu generieren, der nicht alle davon bewertet, sondern nur diejenigen, die zum Ziel führen. Auf die Problematik des Speicherns von längerfristigen Strategien in Zusammenhang mit Entscheidungsbäumen weist Segaran hin (2008, S. 302). Damit ist freilich noch keine Lösung für das Problem gefunden, denn der Faktor Zeit spielt ebenfalls eine nicht unwesentliche Rolle. Stünde unendlich viel Zeit und Speicher zur Verfügung, könnte eine Schach-KI jedes Schachspiel gewinnen.

Der Schlüssel zu einer Maschine, die Schach in akzeptabler Zeit spielt, ist also ein Zuggenerator im Sinne obigen Entscheidungsbaumes, der nach der MinMax-Methode arbeitet oder der verbesserten Alpha Beta-Suche, d.h. also Bewertungen der Stellungen vornimmt, die entweder mehr oder weniger vielversprechende Züge liefern. Wichtig für einen Zuggenerator ist hier, wie tief dieser rechnet. Ergibt sich in einigen Zügen ein möglicher Figurengewinn, kann dies – wer Schach spielt weiß dies - trotzdem ein sog. vergiftetes Opfer sein, das man möglicherweise erst einige Züge später bemerken würde (Steinwender, Friedel, 1995, S.58ff). In diesem Fall würde diese Stellung auch nicht mehr gespeichert, sondern gleich verworfen werden.

Schach-Programme bis zum einleitend erwähnten Deep Blue verfügten hier aber entweder noch nicht über die entsprechende Rechenleistung oder scheiterten an Implementierungsfehlern der Zuggeneratoren. Wie einleitend erwähnt wurde, scheiterte streng genommen auch Deep Blue, da es pures Glück war. Der Grund dafür wird im Fehlervermeidungskapitel erläutert.

6.2 Tic Tac Toe und einfache Algorithmen, bzw. Muster

Tic Tac Toe (TTT) ist im Gegensatz zu Schach ein einfaches Spiel mit drei Spielfeldern in drei Reihen. Ziel ist es, abwechselnd einen Kreis oder ein Kreuz zu setzen und drei Kreise, bzw. Kreuze waagrecht, senkrecht oder diagonal zu erzielen. Gerade die Einfachheit prädestiniert TTT für das Lernen anhand von Beispielen. Wie Yonkers in seinem Buch Tic Tac Tome zeigt, können einer entsprechenden KI sämtliche Züge bei TTT als Eröffnungsbibliotheken eingepflegt werden. Bei Schach dagegen funktioniert dies nur bis zu einigen Zügen, da die Berechnungen der weiteren Zugsmöglichkeiten dann schon ins Astronomische steigen. Beschäftigt man sich intensiver mit den möglichen Zügen und Gegenzügen, kommt man zur Erkenntnis, dass durch Symmetrie des TTT-Spielfeldes lediglich drei Anzugsmöglichkeiten existieren, nämlich Eckfeld, Seitenfeld oder Mitte. Konsequenterweise ist es nicht allzu aufwändig, sämtliche sich daraus ergebenden Stellungen bis zu erzwungenem Gewinn durchzugehen, ganz im Gegensatz zu Schach. Alle anderen Züge führen erzwungenermaßen zu Spielerlust, sofern selbst kein Fehler gemacht wird. Damit dies auch nicht der Fall ist, braucht man nur ab dem dritten Zug auf Algorithmen zurückgreifen, hier drei in einer Reihe des Gegners zu verhindern, bzw. selber zu versuchen drei zu erhalten – vielleicht macht der Gegner ja einen Fehler…

Nach Durchspielen sämtlicher und Weglassen der symmetrischen Möglichkeiten, kommt man zum Schluss, es bei TTT mit einem Spiel zu tun zu haben, das bei optimaler Spielführung immer remis endet und lediglich bei einem Fehler des Gegners ein Sieg möglich wird. Die festgestellte Strategie lautet wie folgt:

1. Anzug: Eck → Gegenzug: Mitte → remis, alle anderen Züge führen zum Spielverlust da drei in einer Reihe erzwungen werden können

2. Anzug: Seite → Gegenzug angrenzendes Eck, Mitte oder gegenüberliegende Seite (Form eines Hummers), alle anderen Züge führen zum Spielverlust, da drei in einer Reihe erzwungen werden können

3. Anzug: Mitte → Gegenzug: Eck, alle anderen Züge führen zum Spielverlust, da drei in einer Reihe erzwungen werden können (Yonkers, 2014, S.1ff.)

Beiden Spielen liegen unterschiedliche Lernstrategien zugrunde. Schach muss sich rein auf Algorithmen verlassen, TTT dagegen kann durch Bilder und Beispiele Muster erfassen und dadurch lernen, nebst obiger durch Algorithmus einprogrammierter Regeln, sich im weiteren Spielverlauf nur darauf zu besinnen, drei in einer Reihe zu verhindern, was in Abweichung obiger Regeln immer zu remis führt. Im Anhang findet sich ein „Schummelzettel“, der es ermöglicht, jedes TTT-Spiel zu gewinnen, oder in Anwendung des oben Gesagten zumindest remis zu spielen. Somit ist hier eigentlich kein Denken erforderlich, sondern algorithmisches Abarbeiten.

Bei Schach dagegen handelt es sich um die Nachbildung eines Denkprozesses. Anders ausgedrückt liegen bei Schach variable Spielmöglichkeiten vor, während bei TTT solche faktisch nicht existieren. Wie o.a. wurde, sind zwar auch die Schachstellungen begrenzt, allerdings realistisch betrachtet nicht zu erfassen. Würde man TTT dahingehend erweitern, dass das Spielfeld und die Anzahl der Reihen größer werden, würde auch hier die Komplexität entsprechend ansteigen. Einfache Muster und Anzugs-Reaktionsregeln reichen dann auch hier nicht mehr aus. Damit ist zum Ausdruck gebracht, je komplexer, im Sinne von umfassender ein Problem wird, desto mehr entfernt man sich von Mustern und Beispielen hin zu strategischer Entscheidungsfindung durch Entscheidungsbäume.

6.3 AlphaGo und Deep Learning

Genau an dieser Stelle wäre Go zu nennen. Das japanische Spiel Go ist bei einem 19x19-Brett noch komplexer als Schach. Im Kern geht es darum, sich Räume zu schaffen und dabei gegnerische Steine zu fesseln. Da die Spielfeldgröße von Go hier theoretsich aber unbegrenzt ist, sind auch die Spielmöglichkeiten unbegrenzt. Klassische Entscheidungsbäume alleine reichen hier nicht aus, um eine Strategie zu entwickeln. Es müssen daher o.a. Bäume mit der Erfahrung aus reiner Brute Force-Rechenleistung kombiniert werden. Die Strategie dahinter läuft dabei wie folgt ab:

„

- Beginnend mit der Ursprungssituation werden solange Spielzüge mit einem zufallsbasierten Verfahren ausgewählt und angewandt, bis eine terminale Position erreicht ist.
- Diese Position kann leicht evaluiert werden, indem der Sieger bestimmt wird.
- Die ersten beiden Schritte werden wiederholt und schließlich derjenige Zug zurückgegeben, der im Mittel am besten abschneidet.

“ (Wäber, 2010, S.7)¨

Führen somit vollständig durchgespielte Partien zum Gewinn, so werden diese Züge in die engere Wahl gezogen. Tritt dagegen Verlust ein, werden diese Züge verworfen. Bezogen auf die im Rahmen dieser Abhandlung von Interesse stehenden Lernstrategien heißt dies, hier einen anpassungsfähigen Algorithmus zu verwenden, da das Go-Brett im Vergleich zu Schach ja dynamisch wachsen oder verkleinert werden kann. Schach-KIs dagegen arbeiten immer in Hinblick auf die Beschränkung durch das 8x8-Schachbrett. So umfassend hier die möglichen Stellungen für Schach auch sein mögen, die Zugmöglichkeiten der einzelnen Spielfiguren sind begrenzt. Anders ausgedrückt gibt es bei Go mit einem 2x2-Spielfeld durch Platzierung der Steine in den Schnittkanten und sämtlichen Zugmöglichkeiten 386.356.909.593 mögliche Spiele wie Tromp mittels seiner quelloffenen Software herausgefunden hat (Tromp, 2016). Bemerkenswert daran ist, dass hinter AlphaGo, also der KI, die letztlich den Weltmeister bezwang, Google und das Unternehmen DeepMind stecken

(Presse, 2016). DeepMind ist nämlich federführend auf dem Gebiet des Deep Learnings.

Was dabei die Erweiterung des Wissens eines KI-Agenten anbelangt, findet dies durch Wahrnehmung statt. Voraussetzung dafür ist jedoch immer ein schon vorhandenes Grundwissen, welches ihm sein Entwickler implementiert hat. Ist ein Agent somit nicht dahingehend konzipiert aus der Wahrnehmung zu lernen, liegt auch eine fehlende Autonomie vor. Rationale Agenten sollen ja geradezu durch Wahrnehmung lernen, also autonom handeln (Russel, Norvig, 2012, S.66). Diesen Gedanken weiter verfolgt die Strategie des Lernens mittels wahrgenommener Belohnungen durch obiges Deep Learning. Hierbei liegt das Schwergewicht darauf, dass die einzelnen Gewichtungen nicht mehr von Programmierern vorgenommen werden, sondern aufgrund von externen Daten, die mittels einer erlernten Prozedur verarbeitet werden (Schlieter, 2015, S.32). Im entferntesten Sinne kann man hierbei von einer Art Lernen durch Erfahrung sprechen, Erfahrung jedoch, welche die KI selbst macht. Demis Hassabis von DeepMind selbst sieht AlphaGo als eine KI, die denkt wie der Mensch (Bambusch [3], 2016, S.19). Deep Learning selbst funktioniert dabei mit ähnlichen Mechanismen wie das menschliche Gehirn. Im Fokus stehen dabei die Neuronen. Eine KI, die selbstständig Verknüpfungen zwischen diesen Neuronen bildet, arbeitet ganz im Sinne des Deep Learning. Was dabei jedoch herauskommt muss nach anderen Methoden bewertet werden, denn nicht jede Kombination von Vernetzung der Neuronen ist auch sinnvoll (Alpaydin, 2016, S.85ff).

6.4 Neuronale Netze

Neuronale Netze arbeiten wie vorhin beim Deep Learning angeführt genau nach dem Schema des Knüpfens von Verbindungen zwischen Neuronen. Wie gut oder schlecht dies geschieht, kann ergänzend zu obiger Strategie des Deep Learnings in Form von Belohnungen gemessen werden, die reelle Zahlen sein können. Denn für Computer bedeuten Süßigkeiten als Belohnung im Gegensatz zu Kleinkindern nämlich nichts, höhere binäre Gewichtungen dagegen sehr wohl. Das System wird somit eine Zielfunktion, wie z.B. optimale Mustererkennung zu erreichen versuchen (Schlieter, 2015, S.30). Damit liegt eine Art des verstärkten Lernens vor. Anzumerken wäre jedoch, hier die binäre Logik zu beachten. Ein Computer weicht nämlich nicht von seinem vorgegebenen Denkmuster im Sinne elektrischer Schaltzustände ab. Konkret bedeutet dies, für dasselbe Problem mit denselben Ausgangsbedingungen immer zur selben Lösung zu kommen. Anders ausgedrückt können KIs nicht lernen, sich unter gleichen Bedingungen abweichend zu verhalten. Sie „lernen“ also immer auf die gleiche Art und Weise und auch immer das Gleiche – auch mit der Gefahr sich hier festzufahren. Eine Anweisungsfolge:

```
a = 0;
while (true)
      {
      a++;
      a--;
      }
```

würde hier immer den Block in den geschweiften Klammern ausführen und bis in alle Ewigkeit ein Inkrement von Null zu Eins vornehmen, um sogleich im Anschluss wieder ein Dekrement des Wertes Eins hin zu Null durchzuführen. Ein Mensch würde an dieser Stelle sofort die Sinnlosigkeit dieser Routine erkennen und abbrechen. Nicht so eine strikt ablaufende Software. Allenfalls könnte an dieser Stelle ein externes Monitoring durch andere Software feststellen, es hier mit einer Endlosschleife, die keinerlei Output erzeugt, zu tun zu haben, dabei aber Rechen-Ressourcen konsumiert. Diese KI könnte dann einen Abbruch herbeiführen, mehr aber auch nicht, denn an welche Stelle in der Software wieder eingesprungen werden sollte, wüsste sie nicht.

Ein anderes Problem neuronaler Netze ist im Moment der nach wie vor begrenzte Speicherplatz und die Zeit, alles Erlernbare auch zu erlernen. Während das menschliche Gehirn über rund 100 Mrd. Neuronen verfügt, die parallel arbeiten können, mag dies zwar schön und gut sein, sie sind jedoch langsamer als Computer. Computer dagegen arbeiten extrem schnell, derzeit aber noch nicht mit optimalem Parallel Processing. Selbst wenn aktuelle experimentelle CPUs es auf um die 1.000 Kerne bringen, würden nicht Mrd. von Prozessen gleichzeitig ablaufen können. Kurzweil schwebt daher an dieser Stelle konsequenterweise der Bau eines – wie das menschliche Gehirn – parallel arbeitenden Supercomputers vor. Bis 2025 könnte ein solcher um nur 1.000 $ erhältlich sein, der dem menschlichen Gehirn in punkto Parallelverarbeitung ebenbürtig wäre (Kurzweil, 2016, S.170f.).

6.5 Der genetische Ansatz

Ein weiterer Ansatz, wie Maschinen lernen, stellt wie beim Menschen die Genetik dar. Beim genetischen Ansatz werden mehrere Programme zur Lösung einer Aufgabe herangezogen, sei dies ein Spiel oder ein individueller Test. Danach werden die besten Programme aussortiert und ein wenig verändert, um noch bessere und schnellere Lösungen zu erzielen, d.h. also die nächste Generation wird geschaffen. Man setzt hierbei die genetische Programmierung solange fort, bis eine Abbruchbedingung erfolgt, die z.B. sein kann:

„
1. die perfekte Lösung ist gefunden
2. eine Lösung, die „gut genug“ ist wurde gefunden
3. die Lösung konnte in vielen Generationen nicht weiter verbessert werden
4. die Anzahl der Generationen hat eine bestimmte Größe erreicht
“ (Segaran, 2008, S.278)

Auch Mainzer erörtert an dieser Stelle die genetische Erzeugung von Systemen mit dem Abschauen aus der Natur. Nicht der Programmierer erzeugt hier eine KI, sondern sie entwickelt sich durch evolutionäre Prozesse basierend auf genetischen Algorithmen. Kritik an der genetischen Programmierung ließ jedoch ebenfalls nicht lange auf sich warten. Von Selektion und natürlicher Auslese war hier die Rede. Will man jedoch Fehler bei KIs vermeiden, muss man jedoch zwangsläufig programmieren wie es in der Natur vorkommt, hat allerdings wie im realen Leben keine Garantie auf die Erreichung des jeweiligen Zieles (Mainzer, 2016, S.93f.).

Anzumerken ist ferner, beim genetischen Ansatz außer Acht zu lassen, bei Modifikationen von Generation zu Generation potentiell neben Verbesserungen auch neue Fehler zu produzieren. Genau wie bei Leben aus Fleisch und Blut können Zellmutationen zu Gendefekten führen. Bei Programmen spricht man dagegen von neu geschaffenen Bugs. In Zeiten der ausschließlichen Assembler-Programmierung war es noch gebräuchlich, selbstmodifizierenden Code zu schreiben. Speicherplatz war zu jener Zeit nämlich knapp, um für ähnliche Programm-Module auch entsprechend viele Routinen im Speicher zu halten. Daher wurden kürzere Modifikationsroutinen entwickelt, die entscheidenden Stellen im Maschinencode dahingehend zu modifizierten, hier z.B. anstatt einer Addition eine Subtraktion durchzuführen. Das Ergebnis wäre, hier mit nur einem Modul zur Steuerung eines Roboterarmes auszukommen, indem dieses sowohl Hebe- wie auch Senkbewegungen durchführen könnte.

Diese Selbstmodifikation wäre auch der Ansatz, wie KIs sich selbst modifizieren könnten und damit ihren eigenen Maschinencode schaffen. Anders ausgedrückt handelt es sich dann um KIs, die andere KIs erzeugen können. Um allerdings komplette Problemstellungen zu kodieren, fehlt es heutigen KIs nach wie vor an entsprechender Intelligenz. Wäre dies möglich, müssten große Software-Hersteller nicht tausende von Programmierer beschäftigen, sondern einfach ein abstraktes Problem formulieren und einer KI z.B. mitteilen: „Schreibe einen Druckertreiber neu“. Die Funktionsweise von Standarddruckern hat sich in den letzten Jahren nämlich nicht grundlegend geändert. Die Ansteuerungsroutinen im weit verbreiteten Windows und Linux ebenfalls nicht. Dennoch braucht jedes neue Modell neu entwickelte Steuerungsroutinen. Es wäre doch herrlich, einfach einer KI die entsprechenden Hardware-Spezifikationen zu übergeben und die KI schriebe den Treiber selbst.

Leider funktioniert es nicht so, da Programmierung von banal erscheinenden Dingen bereits einen astronomischen Aufwand nach sich ziehen kann. Um alleine an alte DOS-Systeme zu denken, erforderte die Bildschirmausgabe eines einzelnen Zeichens bereits dessen Kopieren aus dem Hauptspeicher in den Bildschirmspeicher sowie die Kodierung der Farbe und allenfalls des Setzens des Blink-Attributes. Kommen noch Module zur Speicherung von Texten dazu, mussten komplexe Operationen wie Steuerung der Festplattenmotoren, Bewegung der Schreib- / Leseköpfe, Prüfung auf freien Speicherplatz, Eintragung des Dateinamens in ein Dateisystem und Schreiben der Daten erfolgen.

Es ist daher kein Zufall, objektorientierter Programmierung den Vorzug zu geben, abstrahiert sie doch zentrale Operationen wie Eingabe-, Verarbeitung und Ausgabe und ermöglicht den Fokus auf die ausschließliche Lösung des eigentlichen Problems. KIs profitieren somit aus jahrzehntelanger Weiterentwicklung im Bereich der Informatik und den damit geschaffenen Frameworks sowie Datenbanken wie z.B. Java oder .NET, bzw. Oracle, MS-SQL oder Maria-DB. KIs profitieren aber leider nicht nur davon, sondern erben ganz dem objektorientieren Ansatz der Vererbung folgend, von den in diesen Systemen gemachten Fehlern. Wie im Kapitel zur Fehlervermeidung gezeigt wird, können dies logische Fehler sein, die dann zu katastrophalen Folgen führen können.

6.6 Lernen durch Instruktion und druch Beispiele

Vom genetischen Lernansatz zu unterscheiden ist Lernen durch Instruktion sowie Lernen durch Beispiele. Während Lernen durch Instruktion darauf basiert, Problemstellungen abstrakt zu beschreiben und daraus einen Algorithmus abzuleiten, zielt das Lernen durch Beispiele darauf ab, konkrete Szenarien exemplarisch herauszugreifen. Lernen durch Instruktion zielt dabei auf sowohl endliche, wie unendliche Mengen an Input. Beispiele dagegen sind endlicher Natur. Einwenden könnte man an dieser Stelle, dass man einer KI natürlich auch eine unendliche Zahl von Beispiele zeigt, dann jedoch hätte die KI „nichts gelernt“, sondern lediglich im technischen Sinne eine Zuordnung von Wertepaaren vorgenommen à la: wenn Wert x, dann Lösung y, konkret also algorithmisch gehandelt, was nichts anderes hieße, als wieder Lernen durch Instruktion (Savory, 1985, S.160ff.). Festgehalten werden muss auch, bei einer unendlichen Anzahl an Beispielen auch unendlich lange mit dem Lehren beschäftigt zu sein. Im Endeffekt hieße dies, die KI niemals in Aktion setzen zu können. Werden Beispiele dagegen zu allgemein gehalten, besteht die Gefahr spezielle Ausnahmen zu übersehen. Dies ist z.B. beim Empfang von Spam-Mails der Fall, inwieweit hier exemplarische Beispielwörter herangezogen werden, Werbemails auszusortieren. Wie (in)effektiv derartige Bestrebungen sind, wird deutlich wenn man nur nach reinen Wortfolgen filtert. KIs müssten hier auch den Sinn erfassen können, wobei diesem Unterfangen jedoch Grenzen gesetzt sind (Segaran, 2008, S.5).

An Erkenntnis lässt sich aus diesem Kapitel ableiten, es mit unterschiedlichen Lernprozessen für KIs zu tun zu haben. In der Praxis werden jedoch alle zum Einsatz kommen müssen, um optimale KIs zu generieren. Allerdings würde eine KI, die perfekt arbeitet, wenn sie - theoretisch betrachtet - alles Mögliche erlernt hätte und daher keinen Fehler macht, niemand mehr für einen Menschen halten. Damit hätte „die perfekte KI" den Turing-Test nicht mehr bestanden. Programmierer von KIs stecken somit im Dilemma, eine perfekte KI bewusst auch Fehler machen lassen zu müssen. Bezogen auf Hawking, der wie einleitend erwähnt von KIs spricht, die den Menschen auch übertreffen (nach Kalafat, 2014), wäre genau dieser Ansatz richtig. Hier liegt einmal mehr das zentrale Problem begründet: wie perfekt im Sinne von Unfehlbarkeit eine KI sein soll, darf und muss, bzw. v.a. wie perfekt man eine KI überhaupt machen kann. Jedenfalls kann festgehalten werden, Schach-KIs zu den Vertretern des Lernens durch Instruktion zu zählen, während TTT zu den Vertretern des Lernens sowohl durch Instruktion wie auch durch Beispiele gezählt werden kann. AlphaGo hingegen erfordert Deep Learning-Strategien, also das Lernen aus Erfahrungen. Deep Learning selbst kann dabei auch bei Schach und TTT zum Einsatz kommen. Es ist ja geradezu essentiell, je besser eine KI sein soll, dieser auch mehrere Optionen im Sinne von Lernmöglichkeiten zu bieten. Es stellt sich nur die Frage, welchen Aufwand man hier bereit ist zu betreiben, um ein Ziel zu erreichen.

7 Fehlervermeidungsstrategien in KIs

Um das Problem von Fehlern zu verdeutlichen, sei auf eine interessante Stelle in der TV-Serie „Terminator SCC - Die Sarah Connor Chroniken“ hingewiesen. Lieutenant Ellison merkt dort an, wonach es ein Anfang wäre der KI „John Henry“ die zehn Gebote beizubringen. Wäre dem so, würden KI-Systeme damit keinerlei Gefahr mehr für die Menschheit darstellen. Problematisch für die Realität ist allerdings die Gültigkeit der zehn Gebote erstens nur im Christentum und ferner die Umsetzung des Gebotes: „Du sollt nicht töten!“ Eine KI, welches sich daran hält, würde in gemeinhin als legitimer Aufgabenerfüllung angesehener Terrorbekämpfung versagen. Sie darf ja nicht töten, also auch keine Terroristen. An dieser Stelle kommen sog. Constraints ins Spiel. Beim Vorliegen eines Widerspruchs, bzw. widersprüchlicher Eingangsdaten könnte durch sog. konfliktgesteuertes Backjumping die KI in einen Zustand gebracht werden, der zwar nicht rational erscheint, allerdings die Gefahr minimiert, fehlerhafte Ergebnisse zu produzieren (Russel, Norvig, 2012, S.271). Ob das Resultat an dieser Stelle in Einklang mit Völkerrecht steht oder nicht, sei dahingestellt. Hier geht es nämlich nur um die grundsätzliche Vermeidung von Fehlern. Soll die KI also töten, entspricht dies nicht dem ursprünglichen Gebot. Die spannende Frage ist an dieser Stelle, woran man nun die Constraints knüpft, Terroristen zu erkennen. In den Medien hat man mehr als einmal vernommen, dass fälschlicherweise Zivilisten anstatt Terroristen getötet wurden. Die Alternative wäre sohin der gänzliche Verzicht auf KIs, was jedoch nicht mehr praktikabel erscheint angesichts der hohen Anzahl an Geräten, wie im Kapitel zu KIs im Alltag angeführt wurde. Hier hat somit die normative Kraft des Faktischen schon gesiegt. Kampfdrohnen und intelligente Bomben existieren auf der eine und Sicherheitsaufgaben des Staates andererseits, seine Bürger zu schützen. Wie sich allerdings der Staat selbst vor KIs schützen kann, die Amok laufen, ist nicht bekannt. Essentiell wäre dieser Schutz jedoch und wurde bereits 1942 von Isaac Asimov für Roboter in seinen drei Robotergesetzen formuliert:

„

1. Ein Roboter darf kein menschliches Wesen verletzen oder durch Untätigkeit erlauben, dass ein menschliches Wesen zu Schaden kommt.

2. Ein Roboter muss den von einem Menschen gegebenen Befehlen gehorchen, außer wenn derartige Befehle mit dem ersten Gesetz kollidieren würden.

3. Ein Roboter muss seine eigene Existenz schützen, solange ein derartiger Schutz nicht mit dem ersten oder zweiten Gesetz kollidiert.“

“ (nach Russel, Norvig, 2012, S.1197)

Somit implizieren Asimovs Robotergesetze auch die vorhin angesprochenen Constraints. Bezüglich des Tötens zu Gesetz eins ist aber fraglich, was passieren würde, wenn die Wahl darin bestünde, ein Lebewesen zu töten, um viele

andere zu retten. Bezüglich Gesetz zwei stellt sich die Frage, wie denn eine binär kodierte KI sich generell anderweitig verhalten soll und v.a. wie denn die Schnittstellen zwischen Mensch und Roboter zuverlässig funktionieren sollen, wenn es schon Probleme bereitet, für KIs einen Verstand zu entwickeln, wie im Kapitel zur Entstehung einer KI und im Kapitel zu Lernstrategien erläutert wurde. Letztlich wäre es interessant, wie sich ein Roboter verhalten würde, wenn er von einer KI die Instruktion bekäme, menschliche Infrastruktur zu zerstören - immerhin käme dadurch nicht unmittelbar ein Mensch zu Schaden. Einwenden könnte man hier jederzeit, nur von der jeweiligen Steuerung und Implementierung abhängig zu sein. Es besteht somit auch hier die ständige Gratwanderung zwischen den einleitend erwähnten hart kodierten Robotern und frei programmierbaren KI-Steuerungen. Da Programmierung jedoch ebenso Fehlern unterliegt wie hart kodierte Schaltlogik, besteht in beiden Fällen die Gefahr für Menschenleben. Erinnert sei auch hier an den einleitend erwähnten Schweißroboter, der eine Frau tötete.

7.1 Software und Algorithmen als Basis für KIs

Wie bereits im Kapitel zur Entstehung von KIs angeführt wurde, handelt es sich bei diesen um sog. Agenten, die mittels Aktuatoren entsprechende Aktionen setzen. Die Agenten selbst stellen dabei Software und Algorithmen dar, die immer noch von Menschen entwickelt werden. Da Menschen jedoch Fehler machen, ist es nur logisch, diese Fehler auch in den von ihnen hergestellten KIs vorzufinden. Zumindest einige dieser Fehler lassen sich durch klassische Methoden in der Software-Entwicklung verhindern. Allen voran wäre gründliches Testen zu nennen. Leider kann genau diesem Punkt in immer komplexer werdenden Systemen immer weniger entsprochen werden. Markt- und Konkurrenzdruck zwingen Hersteller dazu, ihre Systeme dann auszuliefern, wenn die Fehleranzahl auf ein akzeptables Maß gesunken ist. Den Beweis dafür sieht man in den ständig nachgereichten Patches für Software, die Fehler beheben, bzw. in Rückrufaktionen diverser Produkte.

7.2 Drei grundsätzliche Arten von Fehlern

Generell gibt es drei Arten von Programmierfehlern und damit auch drei Arten von Fehlern in KIs. Die Art Fehler, die dabei keine Konsequenzen nach sich ziehen, da sie bereits im Vorfeld der Erstellung von Software auftreten, sind Kompilierungsfehler. Ein im Quellcode vorliegendes Programm kann dabei gar nicht erst in für CPUs ausführbaren Code übersetzt werden. Hier wäre v.a. an syntaktische Fehler zu denken, aber auch an fehlende Programmbibliotheken oder falsche Einstellungen für Zielplattformen, wie z.B. fehlende Co-Prozessoren oder Grafikhardware. Da das Programm erst gar nicht übersetzt wird, kann es folglich auch nicht ablaufen und daher auch keinerlei Schaden anrichten, mit Ausnahme von z.B. Entwicklungskosten, Zeitverlusten und damit verbundenem Aufwand für die Fehlersuche im Code sowie Kosten allfälliger Terminüberschreitungen für die verzögerte Fertigstellung.

Davon zu unterscheiden sind die eigentlich kritischen Fehler. Hier geht es primär um Laufzeitfehler. Während das Programm läuft, können etliche Ereignisse eintreten, die zu unvorhergesehenen Zuständen führen und damit zum Programmabsturz- oder -abbruch. Bis zu einem gewissen Grad kann man sich vor Laufzeitfehlern durch Fehlerbehandlungsroutinen noch schützen. An Laufzeitfehlern wäre hier an nicht verfügbare Netzwerke zu denken, weil sich z.B. ein Tablet außerhalb der WLAN-Reichweite befindet, während eine Datenübertragung läuft. Auch volle Speicher wären denkbar, die Entfernung von Speicherkarten, während auf diese zugegriffen wird, etc… Selbstverständlich gibt es auch nicht behebbare Laufzeitfehler, wie z.B. Divisionen durch Null, wenn falsche Werte eingegeben werden oder Situationen eintreten, in denen solche Werte generiert werden.

Zur letzten Kategorie Fehler gehören die kritischsten überhaupt - nämlich die logischen. Diese sind im Programmcode selbst implementiert und spiegeln das Fehlverhalten von Menschen treffend wider (Microsoft, 2007). Ein Beispiel für einen solchen Logikfehler wäre es z.B., wenn ein Warenwirtschaftsprogramm die österreichische Umsatzsteuer von 20% berechnen soll, diesem jedoch die Formel dazu fälschlicherweise als Preis durch fünf anstatt Preis durch sechs eingegeben wird. So mag zwar aus technischer Sicht alles in Ordnung aus logischer Sicht jedoch die Katastrophe vorprogrammiert sein. Der logische Fehler ist hier nämlich, dass im Endpreis bereits 20% Ust. enthalten sind und nicht erst aus dem Endpreis berechnet werden müssen. Der Endpreis liegt daher bei 120% und nicht bei 100%, daher Division durch sechs.

Ein interessantes Beispiel für einen logischen Fehler liefert ferner ein autonomes Fahrzeug betreffend Sperrlinien im Straßenverkehr. Die KI des Fahrzeuges wurde dahingehend programmiert, diese Linien niemals zu überfahren. Offenbar hat man aber auf o.a. Constraints vergessen, diese Regel außer Kraft zu setzen, wenn kein Hindernis im Weg steht. Im Ergebnis wurde die KI dahingehend überlistet, das autonom fahrende Fahrzeug in einem Kreis gefangen zu halten. Das Fahrzeug hätte, wie folgende Abbildung zeigt, gegen die Programmierung verstoßen, den Sperrlinienkreis, den ein Künstler mittels aufgestreutem Mehl angefertigt hat, zu verlassen.

Abbildung 2: „gefangenes“ Auto (Stepanek, 2013)

Hier wird deutlich, dass autonomes Fahren derzeit noch nicht ausgereift ist. Jeder menschliche Intellekt hätte an dieser Stelle begriffen, dass man die Sperrlinie „Mehlkreis“ überfahren darf. Die Frage ist allerdings, wie bringt man einem Fahrzeug bei, was ein Kreis ist und was eine tatsächlich nicht überfahrbare Sperrlinie ist, wo zudem möglicherweise Kinder überfahren werden würden? Die Frage ist bisher ungelöst.

Bezüglich Kreises mag zudem ein Beispiel die Problematik von Fehlern in KIs verdeutlichen. Es stellt nämlich überhaupt kein Problem dar, einen Algorithmus zu entwickeln, der die Kreiszahl Pi bis auf eine gewünschte Stelle errechnet. Ab der wievielten Stelle es jedoch noch sinnvoll wäre, weitere Stellen zu berechnen, ist dagegen schwierig bis unmöglich einer KI beizubringen (Bostrom, 2014, S.152).

Fehlende Stoppregeln können sich ebenfalls fatal auswirken. Wird einer KI nicht mitgeteilt, wann das Ziel z.B. Büroklammern herzustellen erreicht ist, wird diese nicht aufhören diese herzustellen. Teilt man es ihr jedoch mit, würde die KI das Endziel überprüfen, ob dieses erreicht ist. Da kein Fehler gemacht werden darf, würde die KI nicht aufhören festzustellen, ob auch Büroklammern hergestellt wurden, was ebenfalls keine Abschaltbedingung wäre (Ebd., 2014, S.177f). Der Fehler, den Bostrom an dieser Stelle jedoch begeht, ist, dass Maschinen nicht bewusst handeln und beim Erreichen einer zahlenmäßig determinierten Stoppbedingung dieser auch Folge leisten. Alles andere würde auch keinen Sinn machen, da es der Programmierung widersprechen würde. Stoppregeln werden allerdings oft als Sicherheitsabbruchmechanismus in Schleifen verwendet, wenn diese schnell zur Endlosschleife werden könnten.

7.3 Das Kurzschlussverfahren

Eine andere Quelle vieler Fehler, die kaum auffindbar sind, bezieht sich in technischer Sicht auf das sog. Kurzschlussverfahren. Damit ist gemeint, dass in verknüpften Bedingungen ein Compiler den resultierenden Programmcode dahingehend optimiert, nachfolgende Bedingungen gar nicht erst auszuwerten, wenn sich am Gesamtergebnis nichts mehr ändert. So liefert ein binäres ODER bereits dann ein positives Ergebnis zurück, wenn der erste Ausdruck logisch wahr, also eins ist. Tödlich für das Programm wird es allerdings dann, wenn in den folgenden Ausdrücken noch Zuweisungen zu Variablen erfolgen, bzw. sonstige Funktionsaufrufe erfolgen, die dann natürlich nicht mehr getätigt werden würden. Das Kurzschlussverfahren bedeutet somit nicht etwa einen elektrischen Kurzschluss, sondern das Schließen auf ein Ergebnis in kürzerer Ausführungszeit, was grundsätzlich der Laufzeitoptimierung von Programmen dient. Fehler an dieser Stelle zu vermeiden würde allerdings bedeuten, eben auf Kosten gesteigerter Ausführungszeit, das Kurzschlussverfahren zu deaktivieren oder schlimmer noch, den Software-Code zu reengineeren. Befindet man sich in einer derartigen Situation sollte man sich allerdings ohnehin fragen, ob man

denn das Programm ordentlich bei der Entwicklung durchgeplant hat. Kurzschlussverfahrensfehler zählen hier sowohl zu den potentiellen Laufzeit-, wie auch logischen Fehlern bei der Programmerstellung.

```
If ( (a == true)|| (b-- == 0) ) {...}
```

Obiges Dekrement von b würde im Kurzschlussverfahren niemals ausgeführt werden, wenn a den Wert True hat. Wenn der weitere Programmverlauf jedoch auf das Dekrement von b angewiesen wäre und es nicht sicher ist, ob dieses auch durchgeführt wird, ist auch der Programmablauf nicht mehr zuverlässig. Die Gefahr von Endlosschleifen oder falscher Berechnungen steigt hier natürlich immens und damit auch die Gefahr von Fehlern in durch derartige Konstrukte gesteuerter KIs. In der weit verbreiteten Programmiersprache C++ ist die Kurzschlussauswertung jedoch Standard (Kaiser, 2009, S.118)

7.4 Falsche Datentypen und Unterläufe

Fehler entstehen auch durch falsche Anwendung von Datentypen, die dann unerwartete Auswirkungen nach sich ziehen, weil man sie so nicht vorhergesehen hat. So trug sich der einleitende Fall im Spielcasino zu, wonach eine nur positive Werte (oder Null) annehmende Variable eigentlich negativ hätte werden sollen. Die Programmierung war an dieser Stelle völlig korrekt, der Wert der Variablen wurde vermindert, allerdings trat trotz korrektem Programmablauf ein fehlerhafter und nicht vorhergesehener Zustand ein. Hier sollten dem vermeintlichen Glückspilz fast 43 Mio. Euro ausbezahlt werden. Letztlich einigte man sich auf eine Million Euro, was angesichts der Tatsache eigentlich verloren zu haben, eine stattliche Gewinnsumme ist (Jelenko, 2012). Aus technischer Sicht war das Problem dabei ein sog. Über-, bzw. im obigen Casino-Fall ein Unterlauf einer Variablen (Maguire, 1993, S. 150ff.)

7.5 Pufferüberläufe

Mit obigen Variablenunterlaufsfehlern verwandt sind Pufferüberläufe. Hier werden mehr Daten als erwartet eingelesen, die dann Speicherzellen überschreiben. Speicherzellen können dabei sowohl als Werte, Zahlen oder Programmanweisungen interpretiert werden. Egal was hiervon der Fall wäre, die eingelesenen Daten würden dann an dortiger Stelle schlichtweg falsch sein oder noch schlimmer als Programmcode interpretiert werden. Dieser Programmcode kann dann alles Mögliche beinhalten. Normalerweise wird sich aus einer Zufallsfolge zwar kein gültiges Maschineprogramm ergeben, allerdings könnten theoretisch „sinnvolle Sprunganweisungen entstehen“, die dann Routinen ansteuern könnten zur Auslösung des Feuerbefehls in Waffensystemen bis zum Abschuss von Nuklearwaffen. Somit führen derartige Fehler genau zu den Eigendynamiken, die eigentlich nicht stattfinden sollten. Es genügt allerdings schon, wenn Steuerungssysteme an dieser Stelle im Zuge von Denial of Service abstürzen würde oder mit falschen Werten operieren. Handelt es sich dabei z.B. um die Steuerung eines autonom fahrenden LKWs,

die falsche Navigationsdaten empfängt oder Sensoren von der Fahrbahn falsch auswertet, wird deutlich, dass auch im Zivilbereich andere Verkehrsteilnehmer massiv gefährdet werden würden, wenn es zum Unfall kommt.

Hardwaremäßig gibt es hierzu zwar Schutzmechanismen, der Programmcode gegen Überschreiben schützt, was aber dann? Immerhin fehlen dann die Eingangsdaten, die ja geschrieben werden hätten sollen. Auch hier stehen die Chancen „sehr gut", dass das System ohnehin abstürzt. Um also autonom zu Fahren ist zuverlässige Programmierung und extrem viel Rechenleistung nötig. Würde man sich hier verrechnen oder liegen o.a. Fehler vor, so sind Unfälle vorprogrammiert. Konsequenterweise bringt es Danny Shapiro vom Grafikkartenhersteller Nvidia auf den Punkt: „Neben vielen Kameras sind äußerst leistungsfähige und energieeffiziente Prozessoren nötig, die optimiert suchen, natürliche Sprache verarbeiten und Objekte erkennen. Sie müssen interpretieren können, was ein Straßenschild ist, ein Auto, ein Fußgänger, ein Hund oder ein Ball, der auf die Straße rollt." (nach Lingner, 2014, S.106).

Unsinn kam auch beim Absturz der Marssonde Climate Orbiter im Jahr 1999 heraus. Unterschiedliche Wissenschaftler rechneten in unterschiedlichen Maßeinheiten. So wurde einerseits in Metern wie auch Fuß gerechnet. Die Folge war eine falsche Orbitalflugbahn, die zum Absturz in der Mars-Atmosphäre führte (Spiegel, 1999). Hätte man an dieser Stelle bessere Tests durchgeführt und das System dahingehend programmiert, hier auf einheitliche Werte zu prüfen, wäre aller Voraussicht nach nichts passiert. Immer wieder gemachte Fehler, die katastrophale Folgen haben können, sind dabei ja geradezu nicht auf Plausibilität überprüfte Eingangsdaten. Erstens könnten an Stellen, wo Zahlen erwartet werden, alphanumerische Zeichen eingegeben werden, andererseits können Werte angegeben werden, die für ein Programm außerhalb der Spezifikationen liegen, wie eben bei o.a. abgestürzter Sonde. Der Klassiker, der an dieser Stelle noch zu erwähnen wäre, ist der sog. Katzentest, mit dem man einfach irgendwelche Eingaben auf der Tastatur tätigt, so als ob eine Katze darüber laufen würde. Sinnvolles dürfte hier nicht herauskommen (Maguiere, 1999, S.130ff.).

7.6 Das fehlende „Vielleicht" als mögliche Todesursache durch KIs

Keinem Programm ist bisher gelungen, die Entscheidung zu treffen, ob etwas geschehen soll oder nicht, wenn keine eindeutigen Informationen vorliegen. Da es im binären Denken nur Ja oder Nein mit den Zuständen Null und Eins gibt, fehlt auch ein „Vielleicht". Es kann allenfalls nur emuliert werden. Dies könnte wie Alpaydin anführt, durch Erzeugung von Zufallszahlen geschehen (2016, S.32f.). Zufallswerte würden dann als Basis für Verzögerungsschleifen dienen und ein entsprechendes Verhalten emulieren. Allerdings werden auch Zufallszahlen durch Algorithmen mittels diverser Faktoren wie z.B. der aktuellen Uhrzeit, Mausposition und aktueller Systemlaufzeit als Startwert in

elektronischen Systemen erzeugt. Dennoch ließe sich eine Schleife programmieren, die nach einer gewissen Zeit und einer entsprechenden Anzahl an Entscheidungen sich für oder gegen die Ausführung einer Bedingung entscheidet. Grundsätzlich gibt es allerdings keine Entscheidung à la:

- `if ( Muedigkeit = true ) { Wegfahrsperre == true } // wenn z.B. der Fahrer müde ist, ihn nicht fahren lassen`
- `if ( Muedigkeit = false ) { Wegfahrsperre == false } // wenn der Fahrer nicht müde ist, ihn fahren lssen`
- `if ( Muedigkeit = unknown ) { Wegfahrsperre == perhaps } // ??? <-- existiert in programmtechnischer Form nicht - wenn der Status der lenkenden Person unbekannt ist, was dann?`

Eine KI, die an dieser Stelle nicht entscheiden kann, ob eine lenkende Person müde ist oder nicht, kann auch keine Wegfahrsperre zuverlässig freischalten oder nicht. Hier ist die KI auf ihre Sensoren angewiesen. Liefern ihr diese keine eindeutigen Daten, ob die Person müde ist, könnte natürlich ein Fallback auf „Sperre“ erfolgen. Ist die Entscheidung allerdings falsch und die Person putzmunter, wird diese nicht gerade erfreut über die Fehlfunktion sein. Ist es aber umgekehrt, kann es sie das Leben kosten. Hier stellt sich die Frage, ob eine KI töten kann auf eine ganz andere Art, nämlich durch Unterlassung, also ganz im Sinne Asimovs Robotergesetz Nummer eins. Zusätzliche Fragen wären, wie Müdigkeit überhaupt gemessen werden sollte? Entscheidend ist, was die KI machen soll. Klassisch würde man eine Variable mit einer Default-Handlung vorbelegen für den Fall, dass der Status unbekannt ist. Hier könnte die KI dann rückfragen: „Bist du müde?“ Diese Entscheidung trifft dann aber nicht die KI, sondern erstmalig deren Programmierer, was in diesem Fall geschehen soll. Eine andere Möglichkeit wäre, auch hier den Zufall im Sinne von Wahrscheinlichkeit entscheiden zu lassen. Dann aber wäre auch nicht die KI, sondern der Zufallszahlenalgorithmus der Entscheidungsträger. Sicher ist allerdings, in beiden Fällen eine Auswahl auf gut Glück zu treffen.

7.7 Default-Schließen und unsicheres Schließen

Erinnert sei bezüglich Default-Schließens einmal mehr an den einleitend erwähnten Fallback-Zug von Deep Blue, der genau dies durchführte. Denkbar wäre auch als Default-Handlung hinsichtlich obiger Wegfahrsperre die Sperre freizugeben, da die Person im Fahrzeug ist und nicht schläft. Daher ist die wahrscheinlichste Folge, dass sie nicht müde ist. Russel und Norvig führen treffend an, dass diese Art des Default-Schließens nicht bis zu einem gewissen Grad geglaubt wird, sondern sofort geglaubt wird, bis triftige Gründe bestehen, das Gegenteil anzunehmen. Der andere Fall wäre das unsichere Schließen. Die Problematik dabei ist jedoch, dass elektronische KIs binär numerisch denken, während der Mensch qualitativ denkt (Russel, Norvig, 2012, S.641). Jemand der gähnt, muss nicht notwendigerweise auch tatsächlich so müde sein, um ein Fahrzeug nicht mehr lenken zu können. Eine numerisch denkende KI, die Gähnen sofort mit Müde assoziiert, würde hier eine Fehlentscheidung treffen. Somit müssen mehrere Faktoren zusammenkommen, wie z.B. gesenkter Kopf,

länger zufallende Augenlieder, etc... Erst kürzlich war ein TV-Werbespot eines Autoherstellers zu sehen, der diese Müdigkeitserkennung beworben hat. Was aber, wenn diese Müdigkeitserkennung versagt und Menschen dabei getötet werden, weil eben der numerische Schwellenwert zu hoch eingestellt war? Menschen sind eben nicht identisch, wie identisch ausgestattete Maschinen. Daher reagieren sie auch individuell und können auch Entscheidungen, wann sie müde sind, oder nicht, nicht blindlings einer KI anvertrauen. Nach wie vor ungeklärt sind Haftungsfragen, wenn durch Roboter, also auch autonom fahrende Fahrzeuge, Menschen zu Schaden kommen. Nach Isaac Asimovs Robotergesetz Nr. 1 darf ein Roboter einem Menschen niemals Schaden zufügen weder durch eine Handlung, noch durch Unterlassung. Bei einem Unfall mit einem autonomen Fahrzeug passiert allerdings genau das (Heinzelmann, 2015, S.18ff.). Was das Schließen anbelangt, steht man immer wieder auch vor der Herausforderung, mit vernünftigen Schätzungen zu beginnen. Während das Schätzen für einen Menschen kein Problem darstellt, müssen sich KIs einzig und allein auf mathematische Wahrscheinlichkeiten verlassen. Das Problem dabei ist jedoch eine KI, die immer nur das wahrscheinlichste Szenario annimmt, immer dann versagen wird, wenn der andere Fall eintritt (Segaran, 2008, S. 136f.)

7.8 Perfektionsreduktion zur Fehlerreduktion

Vermeidung von Fehlern besteht grundsätzlich in der Reduktion von Perfektion. Würde man z.B. bei einem Staubsaugerroboter das Ziel „sauberer Boden“ definieren, dürfte sich kein einziges Staubkorn mehr auf diesem befinden. Der Roboter würde sich in einer Endlosschleife verfangen, wenn die Saugleistung nicht reichen würde, um sämtliche Staubkörner aufzusaugen und ihm seine Sensoren hier mitteilen, dass eben noch Staub vorhanden ist. Russel und Norvig vertreten hier die Ansicht, nicht danach vorzugehen, was man in der Umgebung haben will - also Staubfreiheit, sondern wie sich der Roboter verhalten soll, d.h. konkret auf ein durchschnittliches Niveau von Sauberkeit abzuzielen. Hier spielen somit subjektive Sichtweisen eine zentrale Rolle. Letztlich ist es auch eine philosophische Frage, ab wann jemand den Boden als sauber empfindet und wann nicht (Russel, Norvig, 2012, S.63).

7.9 Unterdimensionierte Hardware, Stolperdrähte und Selbstabschaltung von komplexen Systemen

Zwecks Kontrolle von Fehlern bietet sich auch das Hemmen an. Dieses zielt darauf, unterdimensionierte Hardware einzusetzen - also weniger CPU-Leistung oder Speicher einzusetzen, in der eine KI nicht optimal ablaufen kann und daher auch nicht ihr volles Potential ausschöpft. Dieser Ansatz ist natürlich suboptimal, da eine KI ja ihren Nutzen dann auch nicht optimal erfüllen kann. Andere Möglichkeiten wären im Sinne von Stolperdrähten entsprechende Mechanismen einzubauen, die die KI abschaltet, bevor es zur Katastrophe kommt (Bostrom, 2014, S.192f). Ähnliche Mechanismen existieren

bereits in handelsüblicher Hardware, was man als geplante Obsoleszenz bezeichnet, wenn Tintenstrahldrucker beim Erreichen einer gewissen Seitenzahl den Betrieb einstellen. Bezogen auf KIs wäre hier an die Selbstzerstörung zu denken und bewusst platzierte Sicherungssysteme. Wie komplex Systeme bereits geworden sind, verdeutlicht folgende Abbildung, in der die sog. Lines of Code ersichtlich sind, aus denen heute Steuersysteme bestehen. Man merkt, es hier mit Größenordnungen zu tun zu haben, die schlichtweg nicht mehr kontrollierbar sind:

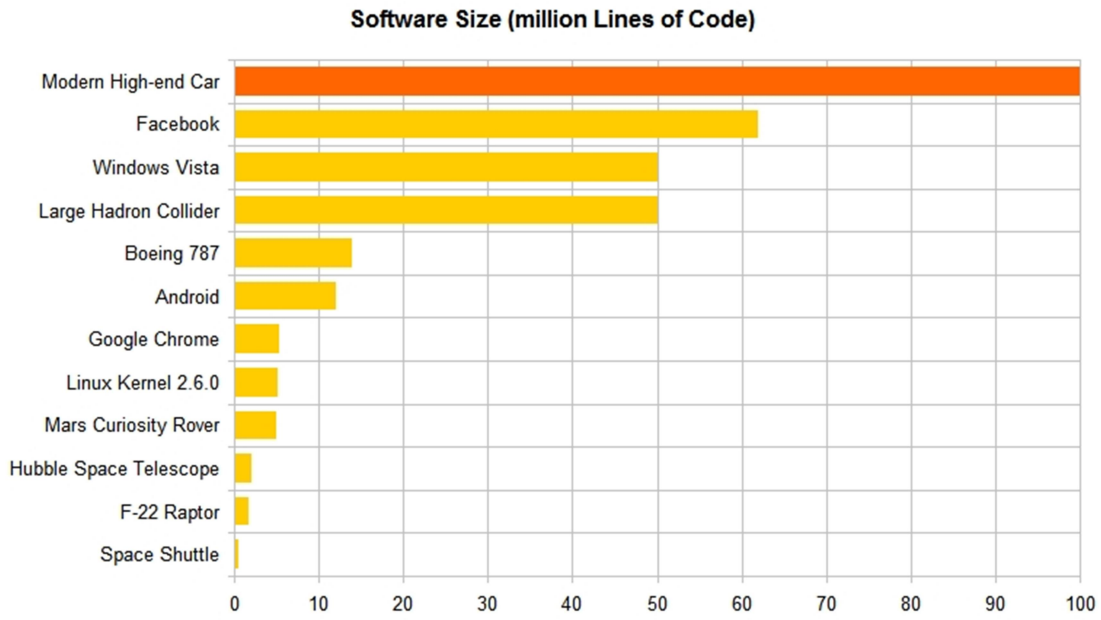


Abbildung 3: Lines of Code in moderner KI (Quelle: Busnelli, 2014)

Dieses Kapitel lässt hier deutlich erkennen, welche Machtlosigkeit Menschen eigentlich vor ihren selbst geschaffenen „Geistern" haben. Fehler können einerseits verheerende Auswirkungen haben, sind andererseits jedoch nicht gänzlich zu verhindern. Bezogen auf Bostroms SI reicht dieser jedoch ein einziger Takeoff, um die Kontrolle zu übernehmen. Im Zeitalter des Internet of Things stehen hier der Menschheit „spannende" Zeiten bevor.

8 Ergebnis und Interpretation

Was KIs sind, ist in der Literatur nicht eindeutig geklärt. Führende Experten vertreten unterschiedliche Auffassungen. Gemeinhin gilt jedoch Alan Turings Aufsatz Machinery and Intelligence von 1950 als der Ursprung der KI-Entwicklung. Das Themengebiet ist als interdisziplinärer Forschungsgegenstand breit gefächert, wo u.a. Psychologie, Religion, aber auch Wirtschaftswissenschaften und natürlich auch Informatik einfließen. Mathematik dient zur logischen Programmierung, Physik zum hardwaremäßigen Aufbau und der Kenntnis um elektrische Schaltungen sowie Chemie zur Anfertigung von Silizium-Chips. Betreffend des Entstehens von KIs muss man sich zudem mit medizinischen Fragen bezüglich Gehirnstrukturen auseinandersetzen. Philosophie beschäftigt sich damit, ab wann von Intelligenz gesprochen werden kann. Wenn man sich mit aktuellen Themen wie dem von Kurzweil prophezeiten Gehirndownload beschäftigt, muss man zum Schluss kommen, es hier tatsächlich mit einer Art Singularität zu tun zu haben, die als KI über allem menschlichen wacht. Entscheidungsprozesse unterliegen dann der Genehmigung oder Ablehnung dieser Singularität. Ob dies die Unsterblichkeit ist, wenn menschliche Existenz in maschineller KI aufgeht, konnte jedoch nicht geklärt werden.

Was jedoch in Zusammenhang mit KIs wichtig ist, ist das Bewusstsein und der menschliche Verstand. Hier konnte eindeutig festgestellt werden, dass Maschinen keinerlei Intelligenz aufweisen – jedenfalls im Moment noch nicht. Wenn allerdings das Szenario wie von Bostrom angesprochen Realität wird, nämlich ausgehend von einer Saat-KI sich die Entwicklung im Zuge einer Intelligenzexplosion zu einer SI vollzieht, bleibt offen, wie sich diese SI dem Menschen gegenüber verhält. Sicher ist jedenfalls, genau diese SI in den Terminator-Filmen als das erklärte Feindbild zu sehen. Von der Hand zu weisen ist dieser Gedanke dabei keineswegs, da führende Experten der IT-Branche, wie Peter Thiel ,Elon Musk oder Bill Gates allesamt düstere Zukunftsszenarien zeichnen, wenn KIs die Kontrolle übernehmen würden.

Geschehen könnte dies dahingehend, wenn eine KI immer mehr und mehr lernt. Gezeigt wurden dabei verschiedene Ansätze zum maschinellen Lernen, wie der genetische Ansatz, wo selektive Auslese der besten Codes erfolgt, das Lernen anhand von Beispielen zur Nachahmung von Verhalten oder Lernen durch Instruktion, wobei dabei mehrheitlich auf Algorithmen gesetzt wird. Festgestellt werden konnte allerdings, dass heutige KIs sehr wohl in der Lage sind einzelne Bereiche zu beherrschen, wie eben Deep Blue das Schachspiel oder AlphaGo Go, allerdings derzeitige KIs nicht über kumulatives Wissen verfügen. Interessant in dieser Hinsicht erscheint jedoch Schwarmintelligenz.

KIs sind jedoch als Produkt aus sog. dualer Forschung sowohl im zivilen wie militärischen Bereich nicht mehr wegzudenken. Der zivile Sektor liefert hier die Forschungen, das Militär sponsert und nutzt die Ergebnisse, wie bezüglich

Google erläutert wurde. In den Medien konnte man sich zudem mehrfach von Kampfdrohnen, intelligenten Bomben und der Entwicklung von Robotersoldaten überzeugen. Problematisch erscheint in diesem Zusammenhang weniger der legitime Einsatzzweck z.B. gegen Terroristen oder zur Abwehr gegnerischer Angriffe, denn die unkontrollierbare Eigendynamik. Diese tritt dann auf, wenn Fehler im Produkt, also der Waffe dazu führen, dass sich diese selbstständig machen könnte und autonom das Feuer eröffnet.

Auch falsche getroffene Ziele zeugen von unzuverlässigen KIs. Da KIs jedoch im Zeitalter des Internet of Things in immer mehr Bereichen vernetzt werden, besteht hier neben BOT-Netzen, die durch feindliche Staaten gesteuert werden auch die Möglichkeit der Übernahme dieser – wie von Hawking einleitend erwähnten – primitiven KIs zum Zwecke der Kontrolle durch eine andere KI selbst. Diese Übernahme von Alltags-KIs ist wie von Antiviren-Experte John McAfee angeführt wurde deswegen möglich, da derartige Geräte schlichtweg bei der Pflege vernachlässigt werden, was deren Wartung mittels Updates betrifft. Damit sind Systeme wie Kühlschränke oder Mikrowellenherde im Sinne der Behebung von Sicherheitslücken gemeint. Wie bei der Schwarmintelligenz gezeigt wurde, lassen sich hier mit dem kollektiven „Verstand" derartiger primitiven KIs sehr wohl größere Ziele erreichen. Ohne Data-Mining zu betreiben wurde hier die Stauprognose durch Mobiltelefone genannt oder die Ausschaltung gegnerischer Infrastruktur durch Mikrowaffen.

Fehlerursachen in Steuerungssoftware, die ja letztlich Kernbestandteil von KIs sind, sind dabei auf drei grundlegende Fehler zurückzuführen. An solchen kommen unwesentliche Compiler-Fehler infrage, sowie die verheerenderen Laufzeitfehler, wenn unvorhergesehene Zustände zur Laufzeit des Programmes eintreten. Katastrophale Auswirkungen dagegen können v.a. logische Fehler nach sich ziehen, die gänzlich unkontrollierbare Abläufe erzeugen. Dazu gehört die Wahl falscher Variablen oder durch obige Fehler verursachte Puffer-Über-, bzw. -unterläufe. Genau dadurch können zufällige Strukturen entstehen, die wichtige Daten oder Programmcode überschrieben. Genau so können die erwähnten Eigendynamiken entstehen, wenn sich hier zufällig gültige Sprungbefehle ergeben sollten, die dann in autonomen Fahrzeugen Unfälle verursachen oder in Waffensystemen den Feuerbefehl erteilen.

Im Endergebnis lässt sich die Gefahr eines real werdenden Terminator-Szenarios daher nicht von sich weisen. Entsprechende Gegenstrategien dafür gibt es allerdings. An diesen wurden etwa genannt, gründlichere Tests, sowie auf die Wahl richtiger Variablen zu achten. Dazu gehört auch das Kurzschlussverfahren zu deaktivieren, bzw. besser auf derartige Konstrukte von Vornherein zu verzichten. KIs auf unterdimensionierter Hardware laufen zu lassen, Stoppbedingungen genau zu prüfen oder eine Art Stolperdraht zu implementieren dient ebenso dazu, hier Sicherungsmechanismen zu implementieren. Im Zivilbereich nennt man dies geplante Obsoleszenz, beim Militär dagegen Sicherheits-Selbstabschaltung oder

Selbstzerstörung. Dies beinhaltet konsequentes Planen der richtigen Verwendung von Variablen, das Überprüfen von Eingangswerten und dem Testen – auch in abgeschotteten Umgebungen, sog. Sandboxen. Auch Stoppregeln müssen beachtet werden, anderenfalls man sich schnell in Endlosschleifen verfängt. Diese Stoppregeln müssen allerdings auch richtig sein. Das gefangen genommene Auto im Kreis wäre hier ein entsprechendes Beispiel dafür. Hiermit ist nicht das stehende Fahrzeug gemeint, sondern die fehlerhafte Stoppbedingung mit fehlenden Constraints à la „Stopp im Kreis, aber nur wenn, kein Hindernis in der Fahrtrichtung ist". Dies könnte man durchaus als Abwandlung von Asimovs Robotergesetzen erachten.

Interessante Interpretationen ergeben sich v.a. vor dem Hintergrund der Kenntnis um die Gefahr fehlerhafter KIs. O.a. führende Experten warnen zwar vor der Übernahme der KIs, stellen derartige Systeme jedoch weiterhin her, forschen in diesem Bereich und nutzen sie auch täglich. Wenn somit deren einhelliger Tenor ist, dass eines Tages KIs die Kontrolle übernehmen und die menschliche Rasse dem Untergang geweiht ist, stellt sich die Frage, warum hier nicht viel mehr gegengesteuert wird. Mögliche Erklärungen können im instinktiven Selbsterhaltungstrieb der Menschen begründet liegen und daher besser bewaffnet sein zu müssen als der andere. Ob dies auf Dauer zielführend sein wird, besonders hinsichtlich der Mikrowaffen, die potentiell feindliche Infrastruktur ausschalten können, sei dahingestellt. Diese können natürlich auch als Bedrohung empfundene Menschen ausschalten. Wie bei der Selbstmodifikation von Code erläutert wurde, können sämtliche Gebote Asimovs oder sonstige Wertvorstellungen der jeweiligen Gesellschaft, die einer KI implementiert wurden, überschrieben werden. Bei Menschen ist als Analogiebeispiel daran zu denken, welche Verhaltensänderungen bei ihnen eintreten, wenn sie dement werden. Bei KIs dagegen gibt es keine Demenz in dem Sinn, sondern eine Umprogrammierung, die jederzeit stattfinden kann. Ein treffendes Beispiel wurde hinsichtlich der Gehirnstimulation genannt, wie Menschen sich hier plötzlich für Superhelden halten, wenn diese am helllichten Tag Autoeinbrüche in aller Öffentlichkeit verüben. Wenn eine KI sich für die Superintelligenz hält, dann wird es aber tatsächlich gefährlich.

Wie im Laufe dieser Abhandlung gezeigt wurde, sind diesbezüglich viele Science-Fiction-Szenarien aus diversen Filmen schon heute Realität geworden. Neben der KI, die wie in Terminator die Menschheit auslöschen möchte, gibt es erste Ansätze von Gedächtnisdownloads von Menschen in Computer, wie in Transcendence oder Battlestar Galactica. Auch wie in Minority Report thematisiert wurde, arbeiten Polizeieinheiten in der Realität mit Verbrechensvorhersagen, bevor diese Geschehen. Da somit überall KIs im Einsatz sind, stellt sich die Frage nicht mehr, ob die Maschinen übernehmen, wie der Titel von Mainzers Buch lautet, sondern man muss zum Ergebnis kommen, dass die Maschinen bereits übernommen haben übernommen haben.

Die offene Frage ist jetzt, wie lange der Mensch die Maschinen noch kontrollieren kann.

9 Literatur

Alpaydin, E. (2016). Machine Learning. Cambridge, London: The MIT Press.

Bambusch, F. (fb) (2015). Kristallkugel mit Big Data. PC Magazin, 2015 (8), 22-25.

Bambusch, F. [2] (fb) (2012). Zukunft ist heute. Wie Roboter und KI unser Leben prägen. PC Magazin, 2012 (8), 100-103.

Bambusch, F. [3] (fb) (2016). Künstliche Intuition. Google-Software bezwingt Go-Meister.. PC Magazin, 2016 (6), 18-20.

Boresch, I., Heinsohn, J., Socher, R. (2007). Wissensverarbeitung. Eine Einführung in die Künstliche Intelligenz für Informatiker und Ingenieure. München: Spektrum Akademischer Verlag.

Bostrom, N. (2014). Superintelligenz. Szenarien einer kommenden Revolution. Berlin: Suhrkamp.

Bostrom [2], N. (2014). Superintelligence. Paths, Dangers, Strategies. Oxford: Oxford University Press.

Busnelli, A. (2014). Car Software: 100M Lines of Code and Counting. Abgerufen von URL: https://www.linkedin.com/pulse/20140626152045-3625632-car-software-100m-lines-of-code-and-counting

Clausen, Jens (2015). Verschwimmende Grenze zwischen Mensch und Technik. Rechtsfragen. Spektrum der Wissenschaft spezial, 2015 (2), 74-79.

Docherty, B. (2012). Losing Humanity. The Case against Killer Robots. New York: o.V.

Dorn, J., Gottlob, G. (1997). Künstliche Intelligenz. In Rechenberg, P. Pomberger, G. (1997), Informatik-Handbuch (819-838). München, Wien: Carl Hanser Verlag.

Eberl, U. (2016). Smarte Maschinen. Wie Künstliche Intelligenz unser Leben verändert. München, Wien: Carl Hanser Verlag.

Heinemann, H. (1986). Einführung in die Industrieroboter-Technik. Allgemeine Einsatzplanung – von der Idee bis zur Realisierung. Essen: Vulkan Verlag.

Heinzelmann, R. (2015). Roboter im Recht. Wer muss haften? PC Magazin, 2015 (12), 16-18.

Hosbach, W. (whs) (2017). Wer spielt denn da? PC-Magazin 2017 (4), 40-43

Hosbach, W. (whs) [2] (2015). Parasit im PC. Botnetze. PC Magazin 2015 (10), 46-49.

Jelenko, M. (2012). Klage nach Gewinn: Casino zahlt 1 Million €. Abgerufen von URL: http://www.heute.at/news/oesterreich/Klage-nach-Gewinn-Casino-zahlt-1-Million;art23655,801180.

Kaiser, R. (2008). C++ mit Microsoft Visual C++ 2008. Heidelberg: Springer.

Kalafat, H. (2014). Physiker warnt vor künstlicher Intelligenz. Abgerufen von http://www.handelsblatt.com/technik/forschung-innovation/stephen-hawking-physiker-warnt-vor-kuenstlicher-intelligenz/11067072.html.

Költzsch, T. (2015). Das Risiko bei KIs ist nicht Bosheit, sondern Fähigkeit. Abgerufen von https://www.golem.de/news/stephen-hawking-das-risiko-bei-kis-ist-nicht-bosheit-sondern-faehigkeit-1510-116815.html.

Kotrba, D. (2017). Freispruch für Tesla: Autopilot nicht schuld an Unfall. Abgerufen von: https://futurezone.at/digital-life/freispruch-fuer-tesla-autopilot-nicht-schuld-an-unfall/242.023.411.

Kronen Zeitung (2017). Roboter tötete Frau: Witwer klagt Firmen. 16.3.2017, 17.

Kurzweil, R. (2016). Die Intelligenz der Evolution. Wenn Mensch und Computer verschmelzen. Köln: KiWi-Taschenbuch.

List, A. (2017). Fliegendes Auge. E-Media, 2017 (03), 30-34.

Lingner, M. (2014) TFT statt Tacho. IT im Auto. PC Magazin 2014 (6), 104-106.

Maguire, S. (1993). Nie wieder Bugs! Die Kunst der fehlerfreien C-Programmierung. Unterschleißheim: Microsoft Press

Mainzer, K. (2016). Künstliche Intelligenz – Wann übernehmen die Maschinen? München: Springer.

Microsoft, Corporation (2007). Erkennen von Fehlern: Drei Arten von Programmierfehlern. Abgerufen von: https://msdn.microsoft.com/de-de/library/s9ek7a19(v=vs.90).aspx.

Münkel, B. (2012). Wie Ameisen denken. Mit Schwarmintelligenz entspannt ans Ziel. PC Magazin, 2012 (5), 114-117.

ORF (2016). McAfee warnt vor vernetzten Geräten. Abgerufen von: http://orf.at/stories/2354091/.

Presse, Die (2016). Googles KI "Alpha Go" schlägt Go-Meister. Abgerufen von: http://diepresse.com/home/techscience/technews/5222925/Googles-KI-Alpha-Go-schlaegt-GoMeister.

Putnik, I. (2017). Also sprach Alexa. E-Media, 2017 (3), 30-34.

Rechenberg, P. (1997). Formale Sprachen und Automaten. In Rechenberg, P. Pomberger, G. (1997), Informatik-Handbuch (77-98). München, Wien: Carl Hanser Verlag.

Russel, S., Norvig, P. (2012). Künstliche Intelligenz. Ein moderner Ansatz. München: Pearson.

Sagaran, T. (2008). Kollektive Intelligenz. analysieren, programmieren & nutzen. Köln: O´Reilly Verlag.

Savory, S. (Hrsg.) (1985). Künstliche Intelligenz und Expertensysteme. München, Wien: R. Oldenburg Verlag.

Schlieter, K. (2015). Die Herrschaftsformel. Wie Künstliche Intelligenz uns berechnet, steuert und unser Leben verändert. Frankfurt: Westend.

Silver, N. (2012). The Signal and the Noise. New York: The Penguin Press.

Spektrum der Wissenschaft spezial (Mai 2015). Mensch, Maschine, Visionen. Wie Biologie und Technik verschmelzen. Heidelberg: Spektrumverlag.

Spiegel Online (1999). Mars Climate Orbiter. Absturz wegen Leichtsinnsfehler beim Rechnen. Aberufen von: http://www.spiegel.de/wissenschaft/mensch/mars-climate-orbiter-absturz-wegen-leichtsinnsfehler-beim-rechnen-a-44777.html.

Standard, Der (2001). Eigene Landkarten für das Humankapital. 10.01.2001, 16.

Steinwender, D., Friedel, F. A. (1995). Schach am PC. Bits und Bytes im könglichen Spiel. München: Markt & Technik.

Stepanek, Martin (2017). Künstler sperrt selbstfahrendes Auto mit Mehl ein. Abgerufen von URL: https://futurezone.at/digital-life/kuenstler-sperrt-selbstfahrendes-auto-mit-mehl-ein/253.925.816.

Strackel, P. (2017). Computer sind dumm. E-Media, 2017 (3), 30-34.

Teahan, W. J. (2010). Artificial Intelligence – Agent Behaviour I. London: Bookboon.

Tuck, J. (2016). Evolution ohne uns. Wird künstliche Intelligenz uns töten? Kulbach: Plassen Verlag.

Turing, Alan (1950). Computing Machinery and Intelligence. Mind, o.D. (49), 433-460

Tromp, J. (2016). Finally Calculated: All the Legal Positions In a 19x19 Game of Go- Abgerufen von: https://science.slashdot.org/story/16/01/24/1428246/finally-calculated-all-the-legal-positions-in-a-19x19-game-of-go.

Wäber, D. (2010). Monte-Carlo Methoden für das Spiel Go. Bachelorarbeit an der Freien Universität Berlin, Berlin.

Wagner, T. (2015). Robokratie. Google, das Silicon Valley und der Mensch als Auslaufmodell. Köln: PapyRossa Verlag.

10 Abbildungsverzeichnis

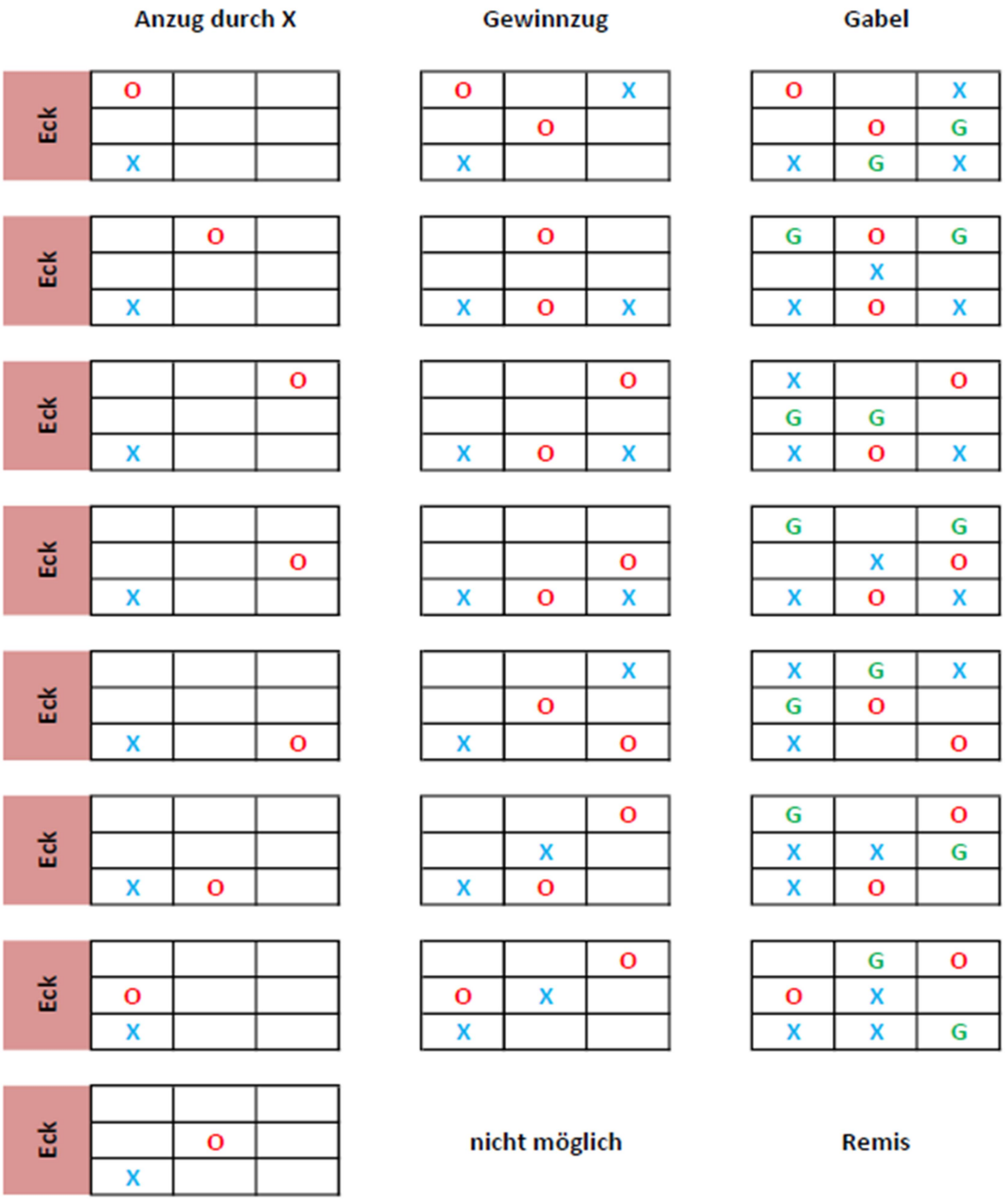
Abkürzungen

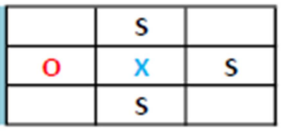
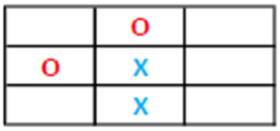
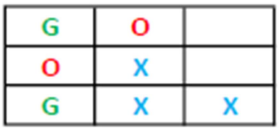
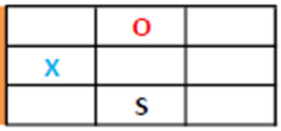
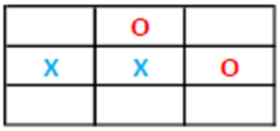
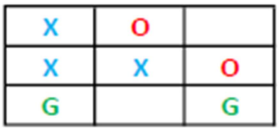
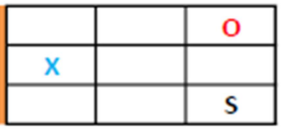
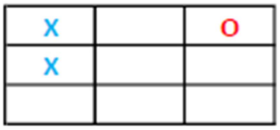
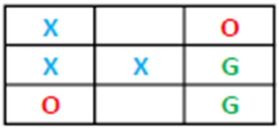
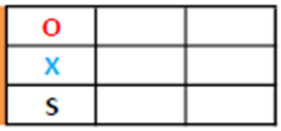
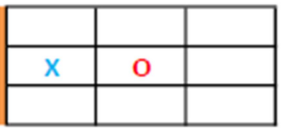
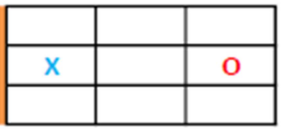
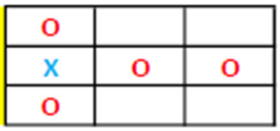
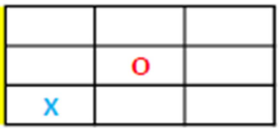
CPU	Central Processing Unit
DB	Datenbank
DI	Diplomingenieur
D.h. / d.h.	Das heißt / das heißt
Ebd.	Ebenda
etc...	et cetera
f.	folgende
ff.	fortfolgende
FH	Fachhochschule
Hrsg.	Herausgeber
KI	Künstliche Intelligenz
KIs	Künstliches Intelligenz-System
LKWs	Lastkraftwagen
MB	Megabyte
Mio.	Million(en)
MIRI	Machine Intelligence Research Institute
MIT	Massachusetts Institute of Technology
Mrd.	Milliarden
o.a.	oben angeführte(m)
S.	Seite
SI	Superintelligenz
sog.	sogenannte(r)
TTT	Tic Tac Toe
u.a.	unter anderem
Ust.	Umsatzsteuer
V.a. /v.a.	Vor allem / vor allem
z.B.	zum Beispiel

Anhang

Im Folgenden finden sich TIC TAC TOE-Gewinn- u. Verteidigungsstrategien (Quelle: Eigenzeichnung mit Excel und grafischer Ausdruck):

X, O = Spielzüge – G = Gable – S = Symmetrie → daher Stellung nur 1x



	Anzug durch X	Gewinnzug	Gabel
Mitte			
Mitte		nicht möglich	remis bei keinem Fehler
Rand			
Rand			
Rand		nicht möglich	Remis
Rand		nicht möglich	Remis
Rand		nicht möglich	Remis
Regel		Anzug Rand, Hummerform	
Regel		Anzug Eck - Mitte	
Regel		Anzug Mitte - Eck	

Eidesstattliche Erklärung

Hiermit erkläre ich ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig angefertigt, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ich erkläre außerdem, dass die vorliegende Arbeit bei keiner anderen Institution (Fachhochschule, Universität, Pädagogische Hochschule oder vergleichbare Bildungseinrichtung) zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum	Unterschrift